

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: June 1, 2003, 22:34:30 ; Search time 5590 Seconds

(without alignments)
4386.849 Million cell updates/sec

Title: US-10-030-330-1
Perfect score: 4438
Sequence: 1 MKKSFLLAVMGFIAMQGH.....PNGAYILKVDGTTKINIVH 843

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

ched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL-frame.p2n.model -DEV-xlh
-Q-cgnt2.1/USPTO_spool/US10030330/runat_23052003_181644_6751/app_query.fasta.1.1031
-DB-GenEmbl -QFAST-fastap -SUFFIX-p2n.rge -MINMATCH-0.1 -LOOPCL-0 -LOOPEXT-0
-UNITS-bits -START-1 -END-1 -MATRIX-blosum62 -TRANS-human40.cdi -LIST-45
-DOCAIGN-200 -THR_SCORE-pct -THR_MAX-100 -THR_MIN-0 -ALIGN-20 -MODE-LOCAL
-OUTFMT-pto -NORM-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-2000000000
-USER-US10030330.ecgn.1.1.3750 -runat_23052003_181644_6751 -NCPU-6 -ICPU-3
-NO_XLPHY -NO_MAP -LARGEORDER -NEG_SCORES-0 -WAIT -DSPBLOCK-100 -LONGLOG
-DEV_TIMEOUT-120 -WARN_TIMEOUT-30 -THREADS-1 -XGAPOP-10 -XGAPEXT-0.5 -FGAPOP-6
-FGAPEXT-7 -YGAPOP-10 -YGAPEXT-0.5 -DELOP-6 -DELEXT-7

Database :

GenEmbl:*
1: gb_da:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_pin:*
35: em_hlg_rod:*
36: em_hlg_mam:*
37: em_hlg_vtl:*
38: em_sy:*
39: em_hlgo_hum:*
40: em_hlgo_mus:*
41: em_hlgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4438	100.0	2532	6	AX048260	AX048260 Sequence
2	1240	27.9	2661	1	POYAAA	M83096 Porphyromon
3	1238	27.9	4056	1	S75942	S75942 prt-99 kda
4	439	11.2	1357	1	STRSPBJ	L26134 Streptococc
5	439	11.2	1357	1	STRSPBJ	L26139 Streptococc
6	439	11.2	1357	1	STRSPBJ	L26133 Streptococc
7	439	11.2	1357	1	STRSPBJ	L26141 Streptococc
8	438	11.2	1479	1	AB051298	L26157 Streptococc
9	436	11.2	1357	1	STRSPBJ	L26154 Streptococc
10	436	11.2	1357	1	STRSPBJ	L26161 Streptococc
11	436	11.2	1357	1	STRSPBJ	L26127 Streptococc
12	436	11.2	1357	1	STRSPBJ	L26133 Streptococc
13	436	11.2	1357	1	STRSPBJ	L26135 Streptococc
14	436	11.2	1357	1	STRSPBJ	L26136 Streptococc
15	436	11.2	1357	1	STRSPBJ	L26137 Streptococc
16	436	11.2	1357	1	STRSPBJ	L26147 Streptococc
17	436	11.2	1357	1	STRSPBJ	L26149 Streptococc
18	435	11.2	1357	1	STRSPBJ	L26156 Streptococc
19	435	11.2	1357	1	STRSPBJ	L26160 Streptococc
20	435	11.2	1357	1	STRSPBJ	L26132 Streptococc
21	435	11.2	1357	1	STRSPBJ	L26140 Streptococc
22	435	11.2	1357	1	STRSPBJ	L26142 Streptococc
23	434	11.1	1357	1	STRSPBJ	L26125 Streptococc
24	434	11.1	1357	1	STRSPBJ	L26151 Streptococc
25	434	11.1	1357	1	STRSPBJ	L26152 Streptococc
26	434	11.1	1357	1	STRSPBJ	L26158 Streptococc
27	434	11.1	1357	1	STRSPBJ	L26159 Streptococc
28	434	11.1	1357	1	STRSPBJ	L26162 Streptococc
29	434	11.1	1357	1	STRSPBJ	L26126 Streptococc
30	434	11.1	1357	1	STRSPBJ	L26129 Streptococc
31	434	11.1	1357	1	STRSPBJ	L26130 Streptococc
32	434	11.1	1357	1	STRSPBJ	L26143 Streptococc
33	434	11.1	1357	1	STRSPBJ	L26144 Streptococc
34	434	11.1	1357	1	STRSPBJ	L26145 Streptococc
35	434	11.1	1357	1	STRSPBJ	L26146 Streptococc
36	434	11.1	1357	1	STRSPBJ	L26148 Streptococc
37	434	11.1	1628	1	STRSPBJ	M86905 Streptococc
38	434	11.1	4334	1	AB030578	AB030578 Streptoc
39	434	11.1	10029	1	AE010112	AE010112 Streptoc
40	434	11.1	53291	1	AE014170	AE014170 Streptoc
41	433	11.1	1357	1	STRSPBJ	L26153 Streptococc
42	433	11.1	1357	1	STRSPBJ	L26155 Streptococc
43	433	11.1	1357	1	STRSPBJ	L26131 Streptococc
44	433	11.1	1357	1	STRSPBJ	L26138 Streptococc
45	433	11.1	1357	1	STRSPBJ	L26150 Streptococc

RESULT 1

ALIGNMENTS

AX048260
 LOCUS AX048260 2532 bp DNA linear PAT 15-DEC-2000
 DEFINITION Sequence 2 from Patent WO0063394.
 ACCESSION AX048260
 VERSION AX048260.1 GI:11877029

KEYWORDS
 SOURCE
 ORGANISM

Porphyromonas gingivalis.
 Porphyromonas gingivalis
 Bacteria; Bacteroidetes; Bacteroides; Bacteroidales;
 Porphyromonadaceae; Porphyromonads.
 1 (bases 1 to 2532)
 TRAVIS, J., POTEMPA, J., and NELSON, D.
 A polypeptide having amidolytic activity for a serpin
 PATENT: WO 0063394-A 2 26-OCT-2000;
 UNIVERSITY OF GEORGIA RESEARCH FOUNDATION, INC. (US) ; TRAVIS,
 JAMES (US) ; POTEMPA, JAN (US) ; NELSON, DANIEL (US)

FEATURES
 source 1.2532
 Location/Qualifiers
 /organism="Porphyromonas gingivalis"
 /db_xref="taxon:837"

COUNT 724 a 561 c 608 g 639 t

Alignment Scores:
 Pred. No.: 5 25e-283 Length: 2532
 Score: 4438.00 Matches: 843
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0

US-10-030-330-1 (1-843) x AX048260 (1-2532)

OY 1 MetLysSerPheLeuLeuAlaIleValMetLeuPheGlyIleAlaMetGlnGlyHis 20
 DB 1 ATGAAAAAAGTTTCTTTTACCCATAGTAATGCTTTGGCATGGCCATGCAAGGACAT 60
 OY 21 SerAlaProValThrLysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgGlnVal 40
 DB 61 TCTGCTCGGTTACGAAAGAGCGACTTGAAGTGGCTGGCTGGCTGGCTGGCTGGCTGG 120
 OY 41 SerLeuArgMetGlyGlnThrAlaValSerAspLysIleSerIleAspTyrValIleArg 60
 DB 121 TCCCTTGGCAATGGGCAACACAGCATCTGCAAGATTTCATCGATTACGTTATTCGG 180
 OY 61 GlnGlyAspAlaGlyIleArgGlyIleThrSerGlnGlnGlySerProAlaTyrPheTyr 80
 DB 181 CAGAGAGATGCTGAGAGGGGATCATCATCAAGAGGAGGCTCTCCATATTTTAT 240
 OY 81 ValAlaAsnArgGlyAsnAsnGlnGlyTyrAlaLeuValAlaAlaAspAspArgIlePro 100
 DB 241 GTAGCTAATCGTGAAATATGAGGGCTATGCTCTTGTGACAGCATGACAGATATACCG 300
 OY 101 ThrIleLeuAlaTyrSerProIleGlyArgPheAspMetAspSerMetProAspAsnLeu 120
 DB 301 ACAATTTTAAAGCCTATTCACCATTTGGCGTTTCGACATGACAGATATGCGGACATATCT 360
 OY 121 ArgMetTyrLeuGlnIleTyrAspGlnGlnIleGlyLeuIleLeuSerGlyLysAlaGln 140
 DB 361 CGCATGTGGCTCAAAATTTACGATCAGAAATAGGCGCTGATCTTCGGAAGAACCTCG 420
 OY 141 LeuAsnGlnGluIleLeuArgThrGlnGlyValProAlaGluValHisAlaLeuMetAsp 160
 DB 421 CTCATATGAAGATATTTACGTACCGAGGGGCTACCGGCTGAGTACATCTCTGATGAT 480
 OY 161 AsnGlyHisPheAlaAsnAspPrometArgTyrAsnGlnGlyTyrProTyrAsnAsnLys 180
 DB 481 AAGGCTCATTTTGGCAACGATCCCATGCGATGAAATCAAGGTTACCCATGCAACATATAG 540
 OY 181 GluProLeuLeuProAsnGlyAsnHisAlaTyrThrGlyCysValAlaThrAlaAlaAla 200
 DB 541 GAACCACTGCTTCTTAATGGCAATCATGCTTATACCGGCTGTGTTGCTACTGCTGACGA 600

OY 201 GlnIleMetArgTyrHisSerTyrProLeuGlnGlnGlySerPheAspTyrHisAla 220
 DB 601 CAATATCATGCGCTTACATAGCTGCGGCTTCAAGGTGAAGGCTTTGATATATCATCA 660
 OY 221 GlySerLeuValGlyAsnTyrPheGlyIleGlnMetTyrAspTyrIleAsnMet 240
 DB 661 GGTTCATATGATGGGCACTGCTGCGGACATTTGGTGAAGTATGACGTGATCAATATG 720
 OY 241 ProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThrLeu 260
 DB 721 CCGGAAATCCCGACCTTGATATCTGACTCAATCTCAAGTGAAGGCTTACGCCACATG 780
 OY 261 MetArgAspValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThrTyr 280
 DB 781 ATGCTGATGATGAGGCACTCTGCTTCGATGATGATTTTGAAGAAAGAGGCTGCTGAC 840
 OY 281 SerValTyrValAlaGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGlnLeu 300
 DB 841 AGCGTTTATGATGAGGAGCCTTGGCAACAACTTCCCTACACGCTTCACTGACGCTA 900
 OY 301 HisValArgAlaLeuTyrThrSerGlnGlnTyrPheAspMetIleArgGlyLeuAla 320
 DB 901 CATGACGCGCTTATATACCTCACAGAGTGGCAGATATGATCCGGGGGAACCTTGC 960
 OY 321 SerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValLys 340
 DB 961 TCCGGAAGGCGGCTTATATGACAGGAAATACAGACATATGACATGCTTCCCTTGGC 1020
 OY 341 AspGlyTyrAlaSerAspGlyThrPheHisPheAsnTyrGlyTyrGlyValSerAsn 360
 DB 1021 GATGGTTATGCTTCCGATGATGATCTTCACTTCACTGAGGCTGGGAGGCTTCCAAC 1080
 OY 361 GlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGlnGlyLe 380
 DB 1081 GCGTCTCAAACTAACACTCTCTCCGCGACTTGTGGATGGAGGTAGGAGATTA 1140
 OY 381 GlyPheThrIleTyrGlnGlnIleIleThrGlyIleGluProAlaLysThrProAlaGlu 400
 DB 1141 GGTTTTACCATTTATCAAGATCATACCGGCTATGCAACCGGCTACCTCCCTGAA 1200
 OY 401 AlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGluAlaGlyTyrLysSer 420
 DB 1201 GCCGATGAGATGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
 OY 421 GlnSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGlyGlnGlnSerAsnLeu 440
 DB 1261 GATCCGATTTGAAGTGAAGGATTCGATATATATATACGAGTGAAGACATCAATCTT 1320
 OY 441 AspLeuGlyTyrArgLeuAsnLysAlaAspGlyGluValIleGluValLysThrSerSer 460
 DB 1321 GACCTCGGATACAGATTGAACAAAGCTGCGAGAGAGTCATAGAGGTGAAACCTTCATCT 1380
 OY 461 IleAsnIleSerTyrTyrGlyTyrGlyGlnHisProGlySerPheSerLeuAlaProAsn 480
 DB 1381 ATCAATATCTTGTGTAGGATGAGTACGAGAGCATCCCGAGAGTGTCTCATGTGCACTAAT 1440
 OY 481 GlnLeuSerGlnGlyIleAsnThrIleThrLeuLeuTyrArgArgThrGlyThrGlnGln 500
 DB 1441 CAGTTGTCAACAAGATCAACACCATCACTTATTCGTCGACAGGACCGCAACAG 1500
 OY 501 TrpGluProValArgHisAlaGlnGlyTyrValAsnSerIleLysValAsnThrThr 520
 DB 1501 TGGAGCGGCTACGCGCATGCAAGGAGGATTTGTCATATGATTAAGTAAATCGACA 1560
 OY 521 AspProAsnAsnValValValThrValAspAsnAsnGlnGlyLysLeuSerIleValPro 540
 DB 1561 GACCCGAACATGTGTGATGTCAGGTAGATTAATACGAGGAGCACTCATGCTCCCC 1620
 OY 541 AsnSerPheValAlaAspLeuAsnSerTyrGlnHisSerThrIleThrValGlnPheAsn 560
 DB 1621 AACAGCTTGTGCGAGATCTGATTTATGAACTATGACATATGAGTATACAGTATCAAT 1680
 OY 561 SerAspSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrGlyAlaThr 580

Db 1681 AGCGACAGCCCTATAGATCCGTACACCCGCTTCTCTATCTACAGAGCTACT 1740
 QY 581 AIAAspValIleSerLeuGlyTyrPvalMetAlaGluValProGlyGlySerSerAsn 600
 Db 1741 GCGGACGATGATATCTTGGGCTGGGTAAATGGCTGAAGTTCCGGCGGTACAGCAAC 1800
 QY 601 TYRProValIleTrpSerLysAspValIleuThrLeuSerGluLysAspTyrThrLeuTrp 620
 Db 1801 TATCCGGTGGTGGTCTTAAAGACGTTCTACCTCTCGGAGGCGCATATACATTTGG 1860
 QY 621 TYRArgPheSerIleAsnAsnGlnLysAspGluTyrLysIleGlySerValSerVal 640
 Db 1861 TATAGATTTTCCATCAACAACAAAGATGATGAAAAAGATCCGAACCGTCACTA 1920
 QY 641 LysThrProThrGluTyrThrHisProLeuPheGluValGlyHisAsnGlnThrSerThr 660
 Db 1921 AAACACCCACAGAGATACGCCCTTATTCGAAAGTGGCCATATCAAACTTCTACC 1980
 QY 661 TYRThrLeuAspMetAlaHisAsnArgValIleuProAspPheThrLeuLysAsnLeuGly 680
 Db 1981 TATACCTGGATATGGACACACACAGATATGGCCGACTTACACTCAAAAATCTCGCA 2040
 QY 681 LeuProPheAsnGlyLysLeuValValIlePheArgGlnThrGlnSerSerSerGlySer 700
 Db 2041 TTGCCCTTCAATGGTGAAGTGGTGTGGTTTCCGCCAAACAAATCTCATGGGGTCT 2100
 QY 701 LeuThrPalaIleGlnGluThrValHisIleLysGlnGlyGluThrPheValTyrLysPro 720
 Db 2101 TTATGGGCGACCTCAACAAACAGTACATATCAACAGAGAGAACTTCCGATATTAACCT 2160
 QY 721 ValIleValGluGlyProIleProAspGlySerTyrArgAlaThrIleuHisAlaPheValAsn 740
 Db 2161 GTTGTGGAAGGCCCTTACTGATGATGATCTTATCGGCGACCTCCATGATCGTAAC 2220
 QY 741 GlyGlnGlnGlnLeuTyrLeuLysGlyLysArgAsnTyrThrValLysIleValAsnGly 760
 Db 2221 GGACAAACACAGTGTGACTCAAGGAGAAAGAACTACACGGTGAAGATCGTCAATGGT 2280
 QY 761 ThrAlaValAlaIleGluSerSerGluGluIleArgValPheProAsnProAlaArg 780
 Db 2281 ACAGCGGTAGAACATATGATCTGACAGAGATAGATATTCCTTAATCCGGACCG 2340
 QY 781 AspTyrValGluIleSerAlaProCysIleProGlnGluThrSerIleIleuPheAsp 800
 Db 2341 GATATGTGAAATATCGGACCTTGCAATCCCAAGAAACATCTATCTTTTCGAT 2400
 QY 801 LeuSerGlyLysIleValMetLysAsnSerLeuSerAlaGlyHisGlyArgMetAspVal 820
 Db 2401 CTGTGACGCAAGTGTCTATGAAGATATTATATACGGGGCATGCGCAAGATGATGTC 2460
 QY 821 SerArgLeuProAsnGlyLysAlaTyrIleLeuLysValAspGlyTyrThrThrLysIleAsn 840
 Db 2461 ACCCGACTTCTTAATGGGCTACATCTTAAGGTGATGATATACGAGAAATTAAT 2520
 QY 841 IleValHis 843
 Db 2521 ATAGTGCAC 2529
 RESULT 2
 LOCUS POYAAA 2661 bp DNA linear BCT 22-MAR-1994
 DEFINITION Porphyromonas gingivalis gene sequence.
 ACCESSION M83096
 VERSION M83096.1 GI:463172
 KEYWORDS
 SOURCE Porphyromonas gingivalis DNA.
 ORGANISM Porphyromonas gingivalis
 Bacteria; Bacteroidetes; Bacteroides; Bacteroidales;
 Porphyromonadaceae; Porphyromonas.
 REFERENCE 1 (bases 1 to 2661)
 AUTHORS Ootago, J. and Kuramitsu, H. K.
 TITLE Isolation and characterization of the Porphyromonas gingivalis prtII

gene, coding for protease activity
 Infect. Immun. 61 (1), 117-123 (1993)
 MEDLINE 93114862
 PUBMED 8093357
 COMMENT On Mar 24, 1994 this sequence version replaced gi:150840.
 FEATURES
 source location/Qualifiers
 1..2661
 /organism="Porphyromonas gingivalis"
 /db_xref="taxon:837"
 BASE COUNT 697 a 662 c 646 g 655 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.3e-72 Length: 2661
 Score: 1240.00 Matches: 299
 Percent Similarity: 52.76% Conservative: 160
 Best Local Similarity: 34.37% Mismatches: 352
 Query Match: 27.94% Indels: 59
 DB: 1 Gaps: 22
 US-10-030-330-1 (1-843) x POYAAA (1-2661)
 QY 1 MetLysLysSerPheLeuLeuAlaIleValMetLeuPheGlyIleAlaMetGlnGlnHis 20
 Db 55 ATGAACGATCTCTTACACCTTAGGGCTTATTTACTGTCTCCCTATGCTC---CAG 111
 QY 21 SerAlaProValThrLysGluArgAlaLeuSerLeuAlaArg-----LeuAlaLeuArg 38
 Db 112 GCAGGACCGGTGACAGATCCAAAGCCGAAACAGACGGCTTAAGAACTTTTGGCCAAACGA 171
 QY 39 GlnValSerLeuArgMetGlyGlnThrAlaValSerAspLysLysIleSerIleAspTyrVal 58
 Db 172 CAACCCACGCTG-----TCTTATCGACACCGAGTCTCCGGATGATTCGTT 219
 QY 59 TYRArgGlnGlyAspAlaGluArgGlyIleThrSerGlnGlnGlySerProAlaTyr 78
 Db 220 TACAAAGCTGCACAAAGAGAGAG-----GCACATA 249
 QY 79 PheTyrValAlaAsnArgGlyAsnAsnGlnGlyTyrAlaLeuValAlaAlaAspArg 98
 Db 250 TTCTGCTTTTCAATCGAGAGAGAAAGACGAGATTTCTCTCTCGACGGATGATCGG 309
 QY 99 IleProThrIleLeuAlaTyrSerProIleGlyArgPheAspMetAspSerMetProAsp 118
 Db 310 TTCCCGAGGTGATCGGATATGCTTTCACAGGGGCACTTCATCGCGCCCTATGCCGAC 369
 QY 119 AsnLeuArgMetTrpLeuGlnIleTyrAspGlnGluIleGlyLeuLeuLeuSerGlyLys 138
 Db 370 AATCTCAGGGGGGTGCTCAAAAGCTATGAACGTGAATGCTTGTCTGAATAGACGGCAAG 429
 QY 139 AlaGlnLeuAsnGlnGluIleLeuArg-----ThrGlnGlyValProAlaGluVal 155
 Db 430 GCAGACCGCATGATCTTATCCCTGGAAGCCCAACCTACACAGGAGACCTGCATCATCAT 489
 QY 156 HisAlaLeuMetAspAsnGlnGlyHisPheAlaAsnAspMetArgTrpAsnGlnGlyTyr 175
 Db 490 GCCCTATTTTGGAAACGGGCGAACAATGCAATGCAATGCAATCTTGTGGATCAGGGCTAT 549
 QY 176 ProTyrAsnAsnLysGluProLeuLeuProAsnGlyAsnHisAlaTyrThrGlyCysVal 195
 Db 550 CCATTTAACACCTTGCATCCCTGCTTCCGCGAGAGACGCTTATACGGTGTGCT 609
 QY 196 AlaThrAlaAlaIleGlnIleMetArgTyrHisSerTrpProLeuGlnGlnGlySer 215
 Db 610 GCCACCGCCATGGAGCAATATCATGCTACATCAAGTGGCGGAGAGAGCTTCCGCGAA 669
 QY 216 PheAspTyrHisAlaGlySerLeuValGly-----AsnTrpSerGlyThrPheGly 232
 Db 670 TATGACTACTAC---GATGATATGACGGGAGCCATACCACTATTCACGACATTCGGT 726
 QY 233 GluMetTyrAspTrpIleAsnMetProGlyAsnProAspLeuAsnAsnLeuThrGlnSer 252
 Db 727 GAGACCTATTAACATCGAGCAAGATGCCGGTAAACATCATGCTT---GTTATCACCCCTGAA 783

QY	253	GlInValAspAlaTyrAlaThrLeuMetArgAspValSerAlaSerValSerMetSerPhe	272
Db	784	GAAGTAAAGAGCGCTTAAAGTACTTTCATGAGATGTCACACTTTCGTCATTAATGACAGTTT	843
QY	273	TyrGIuAnGlySerGlyThrThrSerValTyrValValGlyAlaLeuArgAspAsn	292
Db	844	GCCGATTTTCGCGACAGTGGACCTTTCCTCCATATTGTCGAAAGCGCCCTTCGGGAGACCTTT	903
QY	293	ArgTyrLysArgSerLeuGlnLeuHisValArgAlaLeuTyrThrSerGlnUTRPrts	312
Db	904	CATTACAAAGAAATCCCTTCCTGTTACATCCATGCGTCTCTCTTCCCGGTAAAGATGAGAA	963
QY	313	AspMetIleArgGlyGluLeuAlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGln	332
Db	964	GACATGATTCGTAAAGCAACTGGCAGAAAAACACCTCGTATTATGCGGTGACAGACGA	1023
QY	333	SerIleGlyHisAlaPheValCysAspGlyTyrAlaSerAspGlyThrPheHisPheAsn	352
Db	1024	TCGATGGGACATGCTTCTGTTGGCANTGGATACGAAACGACGAGAACTTCACACTTCAAC	1083
QY	353	TyrGlyTyrPglGlyValSerAsnGlyPheTyrLysLeuThrLeuLeuSerProThrSer	372
Db	1084	TGGGATGGGGTGGCATGATGATAGACGTAATTTCTATCTCATCTCATCTCAATCCGGGTTGC	1143
QY	373	LeuGlyIleGlyGlyGlyGlyIleGlyPheThrIleTyrGlnGluIleIleThrGlyIle	392
Db	1144	CTTGGCACAGCGCGACGGATGAGAGCTACTACTGACAAAGAGTTGTCAATAGCATA	1203
QY	393	GluProAlaLysThrProAlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLeuLys	412
Db	1204	GAGCGCGGACGATGAAGCA---CCCGGTATCGTACCGGATCCGACATCAGCTCAT	1260
QY	413	AspIleGluAlaGluTyrLysSerGlnSerGlyLeuAsnValGlyTyrSerIleTyrAsn	432
Db	1261	GGATTGCAA---CAACAATATGTCTGATGATAGACACTGACCTGACGAGCTAAATCAAGAAC	1317
QY	433	ThrGlyGluGluGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyGlu	452
Db	1318	TACTCCACTTATGACGAGGATGTGAATATGGCTTATCCCTGACGCTTCCCAATGAAACG	1377
QY	453	ValIleGluValLysThrSerSerIleAsnIleSerTyrGly---TyrGlyGluHis	471
Db	1378	GAAACGACTAACCCAGCCGTCACCGTCCCTATCTGTTGGAGAACATTATCGGAGAAAGC	1437
QY	472	ProGluSerPheSerLeuAlaProAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeu	491
Db	1438	ACGGGCAACATTAACATACATACATGCATGCTAATTTGCCGAGGGGAAAAATACCATCTCCATA	1499
QY	492	LeuTyrArgArgThrGlyThrGluGlnTyrPgluProValArgHisAlaGlnGlyTyr	511
Db	1498	CTCTATTCACACAGATGAGTGGCAGATTGGAAGAGCTTAAGCAATACTATATGGCGCTT	1557
QY	512	ValAsnSerIleLysValAlaSerThrThrAspProAsnAsnValValThrValAspAsn	531
Db	1558	GTCATTAAGATAGAAGTG---ACTATGCTTCGCCGAGATGTAGCTTACTCGGTT-----	1608
QY	532	AsnGlnGlyLysLeuSerIleValProAsnSerPheValAlaAspLeuAsnSerTyrGlu	551
Db	1609	GCCGATGGCAGAGTGTCTTCAAAGACGGTCTCTTTCGCACACMACCTGAAGCATATTC	1668
QY	552	HisSerThrIleThrValGlnPheAsnSerAspSerProAspGluIleArgThrProVal	571
Db	1669	GACTGCAAAACACAGTGCACACAGTCTAACAAACCCGGGTACGAAAGAAATTTCTAGTACGCGTC	1728
QY	572	AlaPheAlaLeuSerThrGlyAlaThrAlaAspAspValIleSerLeuGlyTyrValMet	591
Db	1729	ACCTTTGGCCCTCCGCAAT---ACAAGGAGCGGCTCTATTCTTGGGACGACATTTA	1782
QY	592	AlaGluVal---ProGlyLysSerSerAsnTyrProValValTyrSerLysAspValLeu	610
Db	1783	GTCAGATTTACACCCGGGAGATGAAAGACGGCGAAAGATTTGGCTGTACACATTAACAGGCTTC	1842

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	REMARK
611	Thl1leuserg1ng1llyserthrleu1pry1ltyg-----PheSer1leAsn1ng1	628											
1843	AAGGCTGCTGCAGGCAAAATAGCTGCTCTGTCAAGGGCAATGATGAAATGCTTATGGA	1902											
629	LysAspGluTrpLys1leGlySerValSerValLysThrProThrGluTyrThr1s	648											
1903	GATGGCTCATGGATAGATAGATGGCTTCCATTAAGTAAAGTAAAGCAATAGCAACCATTA	1962											
649	ProLeuPheGluValG1ly1sAsnG1lThrSerThrTyrThrLeuAspMetAla1sAsn	668											
1963	TGCTTACTGCTGGGCTCCAAACCCAGATCATGATTTCTCAACATGTTCAATCGGGCAATCC	2022											
669	ArgVal1leuProAspPheThrThrLeu1sAsn1leu---G1leuProPheAsn1G1leu	687											
2023	GAGACATTTGCGCACTTTCAGCATTAACAATTAAGAGTGCTGCTACTTCTCCGGGAATTC	2082											
688	Val1Val1PheAspG1lThrG1lSerSerSerGlySer1leuThrAla1leG1l1uthr	707											
2083	GAAATAGTGGCTATTAAGGCTTTCTCGAAACT---TTCCTCCAAAGCAAAAGAAACAC	2139											
708	Val1s1le1le1sG1ng1l1uthrPheVal1TyrLysProVal1AlaG1lGlyPro----	725											
2140	ATGAGTCTCGCCCAAGGGAAACCAAGATATTGTCTCCGAGCTGCACTGCAACTCTTCT	2199											
726	-----1leProAsp1LysSerTyrAlaG1lAlaThr1leu1sAlaPhe	738											
2200	CTCATATCAAAATGCCCAACTCTTCCGATGGCATATTAACATTTGCATGACA-----	2253											
739	ValAsnG1yG1ng1l1eUtyr-----Leu1sG1yLysArgAsnTyrThr	754											
2254	-----GAGCAAGGATTTTGGGATGCCATTCATTTGTTGGGAGCATATTACTATTCGT	2304											
755	ValLys1leVal1AsnG1lThrAlaVal1Ala1leG1lSerSerGlu-----Glu1le	772											
2305	ATCCGCTCATATACGAGATCTATCCCTCTCGGADATCCGCTGTGAAGATGTTTCTACTATA	2364											
773	ArgVal1PheProAspProAlaArgAspTyrAlaG1l1eSerAlaProG1s1leProG1n	792											
2365	GTACTTATGCCAAATCGCTGTCAGACGTCGACAGCTTCATGATGAGCATTTCCACCATATGCC	2424											
793	GluThrSer1le1leLeuPheAspLeuSerG1yLys1leVal1MetLysAsnSerLeu-Se	812											
2425	GCGACGACACTTCCTGTTTGTGATATTAAAGCGCAATGACAGCTCCACAGAAATCCGA	2484											
812	TAlaG1y1sG1yArgMetAspValSerArgLeuProAsnG1yAlaTyr1leLeu1sVa	832											
2485	ATCTCGCATATGCGCTCTCGACGTCGACAGCTTCCTCCGAAGGCAACATATCGTTGGT	2544											
832	1AspG1yTyrThrThrLys1leAsn1le 841												
2545	AGAAACATGTAGGGAAGCTTTTATT 2572												
RESULT 3													
LOCUS	S75942	4056 bp	DNA	linear	BCF 27-JUL-1995								
DEFINITION	prt1-99 kDa cysteine protease/hemagglutinin [Porphyromonas												
ACCESSION	S75942												
VERSION	S75942.1												
KEYWORDS	GI:913136												
SOURCE	Porphyromonas gingivalis ATCC 53977.												
ORGANISM	Porphyromonas gingivalis Bacterioides; Bacteroides; Bacteroidales; Porphyromonadaceae; Porphyromonas.												
REFERENCE	Madden,T.E., Clark,V.L. and Kiranliou,H.K. Revised sequence of the Porphyromonas gingivalis prt1 cysteine protease/hemagglutinin gene: homology with streptococcal pyrogenic exotoxin B/streptococcal proteinase Infect. Immun. 63 (1), 238-247 (1995)			</									

entry [NCBI gi1687] from the original journal article.

This sequence comes from Fig. 2.

Authors indicate that this paper corrects previously published sequence with accession No. M83096.

FEATURES

Location/Qualifiers

1..4056

source

/organism="Porphyromonas gingivalis"

gene

/db_xref="taxon:837"

CDS

766..3426

note="prt"

note="prt"

gene="prt"

/note="99 kDa cysteine protease/hemagglutinin: This sequence comes from Fig. 2; Protein sequence is in conflict with the conceptual translation;

mismatch(11M>11)"

/codon_start=1

/product="prt"

/protein_id="AAB32891.1"

/db_xref="GI:913137"

/translation="MIASSASPAEQKONAMIKRIEYTLGLLCLPMLQAPVTR

KAQYAKNPFARQPTSSSTASLMDFVYKAEREALFEVFNKEDGFLVAD

REPEVIGAFKGFDAKMPDLRGLKGEREMLVADKAEPIPIREAPTRDL

SSIAPILETGEHSDPILMDQGPENTLHPILPSGOQAYGVATAMQIMRHYKPE

KASGEYDYDMGTGHTHYSGFGEYTMKMPGNIISVIGISPEEKALSTPMROVS

VNMORADGSGFSEFYRALRETHYKSLRYIRSLIPKEMKDIKRELAENRPV

YTAGDMSKGFHVDYEPDGFTHFMGWSNGNTLYNLNGSLGTDGSGYS

TDQEVYIIEPASNAPGIVDPPTTLGLDHMSDELADSVKIRKNTSYAGDVDA

YRLPLNGETETNPVAVIWEEDIGESTGNTIPCSQFAGKRTSILYTDMDAL

WKEKHLIMGLVANKIEYMPAGDAVYADARIYVADKSLSHNKATYKATVY

PCTEERFRTVRLNTEGRLYFLGRHLYELHVPDEDEKYSITITGKARAGOMLV

CTGDELMEDASWIELASIEVAERTSHSLIVANPOIDLTTHRANPELTFESI

TNEGATSSKRIEYAIKAFSETPFOAKEEHSIAOGERTKVLSPELANSSITAE

FDGYIYVIRBOGFWDPLDGDYIRIRILITDSSDIAGKDVSTLYLPNPAHY

VHALPPTVAGSTLRLDIOGRMOLSTRIRICRYASRRRTSEGHLYRCGRHGREAL

YSLIOSVNSKDKGNAGWEISGHCPTVC"

BASE COUNT 1119 a 949 c 911 g 1077 t

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

1

Gaps:

22

Length:

Matches:

Conservative:

Mismatches:

Indels:

299

160

352

55

22

4056

299

160

352

55

22

4056

299

1 MetLysLysSerPheLeuLeuAlaLeuAlaMetLeuPheGlyIleAlaMetClnGlyHis 20
820 ATGAACGATCTTACACCTTAGGCGTATTATTACTGTCTCCCTATGCTC---CAG 876
21 SerAlaProValThrLysGluArgAlaLeuSerLeuAlaArg-----LeuAlaLeuArg 38
877 GCAGACCGGTGACGATCCAAACGACGACGCGCTAGAACTTTTTCCTCCAAACA 936
39 GluAlaSerLeuArgMetGlyGlnThrAlaValSerAspLysIleSerIleAspTyrVal 58
937 CAAACCCAGCTTG-----TCTTCATCGACGACGAGCTCTCCGATGATGATTCGTT 984
59 TyrArgGlnGlyAspAlaGluArgGlyIleThrSerGlnGlnGlySerProAlaTyr 78
985 TACAAGCTGCAAGAAAGAGAGAG-----GCACTA 1014
79 PheTyrAlaAlaAsnArgLysAsnAsnGlnGlyTyrAlaLeuValAlaAlaAspArg 98
1015 TTCTTCGTTTCAATCGAGAGAGAAAGAGATTTCTTCCTGTCGACGCGATATGCG 1074
99 IleProThrIleLeuAlaTyrSerProIleGlyArgPheAspMetAspSerMetProAsp 118
1075 TTCCGCGAGGTGATGATATGCTTTCAAGGCGGCGACTTCGATGCGCGGCTATGCGGAGC 1134

119 AsnLeuArgMetTyrLeuGlnIleTyrAspGlnIleGlyLeuIleLeuSerGlyLys 138
1135 AATCTAGAGGGGTGCTCAAGAGCTTGAAGCTTGTCTATGATGAGCGCAAG 1194
139 AlaGlnLeuAsnGlnGluIleLeuArg-----ThrGluGlyValProAlaGluVal 155
1195 GCAGACCGGATGATGCTTCCGTGAGACCAAGCTTACACGGGAGCTGCGATATCCAT 1254
156 HisAlaLeuMetAspAsnGlyHisIlePheAlaAsnAspProMetArgTyrAsnGlnGlyTyr 175
1255 GCCCTATTTTGAAGAGGCGGACATGATGATGATGATGATGATGATGATGATGATGAT 1314
176 ProTyrAsnAsnGlnGluProLeuLeuProAsnGlnGlyHisIleAlaTyrThrGlyVal 195
1315 CCATTTAACACCTTCATCCCTGCTTCCGCGGACAGCGCTTATACCGGTGTCTT 1374
196 AlaThrAlaAlaAlaGlnIleMetArgTyrHisSerTyrProLeuGlnGlyGlySer 215
1375 GCCACCGGATGAGCAAAATCATGCTCTACAAAGTGGCCGAGAAAGCTTCCGCGCA 1434
216 PheAspTyrHisAlaGlySerLeuValGly-----AsnTyrSerGlyTyrPheGly 232
1435 TATGACTACTAC---GATGATATGAGGCGGACGATACCCACTATTCAGACATTCGCT 1491
233 GluMetTyrAspTyrIleAsnMetProGlyAsnProAspLeuAspAsnLeuThrGlnSer 252
1492 GAGACCTTAACTGGACGAAGTGGCCGCTTAACTACATGAGTGT---GATATCCGCTGAA 1548
253 GluValAspAlaTyrAlaThrLeuMetArgAspValSerAlaSerValSerMetSerPhe 272
1549 GAAGTAAAGCTTAAAGTACTTCAAGAGATGACAGTTCGTCGTAATGCACTTT 1608
273 TyrGluAsnGlySerGlyThrTyrSerValTyrValGlyAlaLeuArgAsnAsnPhe 292
1609 GCGGATTCGCGGACGAGGCGCTTCCATATTTGCGAAGCGCCCTCGGAGAGACCTTT 1668
299 ArgTyrLysArgSerLeuGlnIleHisValAlaGlyAlaLeuTyrThrSerGlnIleTyrPheHis 312
1669 CATTCACAGAAATGCTTCTGTTACATCCATCCCTCTCTCTCCGCTGAGGAAAGGAAA 1728
313 AspMetIleArgGlyGluLeuAlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGln 332
1729 GACATGATTCGTAAGAGACTGCGCAAAACACCTGCTATATGCGCGTCACAGCA 1788
333 SerIleGlyHisAlaPheValCysAspGlyTyrAlaSerAspGlyTyrPheHisPheAsn 352
1789 TCGATGGGACATGCTTCTGCTTGCATGATGATGATGATGATGATGATGATGATGATGAT 1848
353 TrrGlyTrrGlyGlyValSerAsnGlyPheTyrLysLeuThrLeuLeuSerProThrHis 372
1849 TGGGATGGGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1908
373 LeuGlyIleGlyGlyGlyGlyIleGlyPheThrIleTyrGlnGlnIleIleThrGlyLe 392
1909 CTTGGACACAGGCGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1968
393 GluProAlaIleTyrProAlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLeuLys 412
1969 GAGCGCGGACGACAAAGAGCA---CCGCTATCGTACCGGAGTCCGACATCAGCTCTAT 2025
413 AspIleGlnAlaGluTyrLysSerGlyLeuAsnValGlyTyrSerIleTyrAsn 432
2026 GGATTCGAA---CACATATGCTGATGACATGATGATGATGATGATGATGATGATGATGAT 2082
433 ThrGlyGlnGlnGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyGlu 452
2083 TACTCCACTTATGACGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2142
453 ValIleGluValLysThrSerSerIleAsnIleSerTyrPyrGly---TyrGlyGluHis 471
2143 GAAACGACTAACCCGCGGCTACCGCTTATGATGATGATGATGATGATGATGATGATGATGAT 2202

Oy	472	ProGluSerPheSerIleAlaProAsnGlnIleuSerGlnGlyIleAsnThrIleThrLeu	491
Db	2203	ACGGGCACACATACCATTACCATGCAGTCAATATTGGCCAGGGGAAAAATACCATCTCCATA	2262
Oy	492	LeuTyraArgaGthrGlyThrGluGlnTrpGluProValaGlnHisAlaGlnGlyGlyThr	511
Db	2263	CTGTATTCAGAACAGATGGCATGGCATGTGGAAAGAGCTTAAGACATATCTATATGGCGCT	2322
Oy	512	ValaAsnSerIleLeuValaAsnThrThrAspProAsnAsnValaValaThrValaAspAsn	531
Db	2323	GTCATATAGATAGAAGTGG--ACATAGCCTCCCGAGAGATGTAGCTTACTCGGTT-----	2373
Oy	532	AsnGluGlyIleuSerIleValaProAsnSerPheValaLeuAspLeuAsnSerGlyGlu	551
Db	2374	GCCCATGCCAGAGATCGTTCTCAAAAGCGCTTCTCTTCCGCAACCTGAACCATATTC	2433
Oy	552	HisSerThrIleThrValaGlnPheAsnSerAspSerProAspGluIleArgThrProVal	571
Db	2434	GATCGACAAATCAGACGACAGTGTACAAACCGGGATCGGAAGAATTTGTAATGCGCGTC	2493
Db	572	AlaPheAlaLeuSerThrGlyAlaThrValaAspAspValaIleSerLeuGlyTrpValMet	591
Oy	2494	ACCTTTGGCCCTCCGCAT-----ACAAAGGAGCGGCTATTTCTCTGGAGACATTGA	2547
Db	592	AlaGluVala----ProGlyGlySerSerAsnTyrrProValaValaTrpSerGlyAspValLeu	610
Db	2548	GTAACATTAACACCCGGAGATGAAGAGCGGCAAAAAGTTGGCTGCACATTAACAGCGCTC	2607
Oy	611	ThrLeuSerGlnGlyAspTrpThrLeuTrpTyraG-----PheSerIleAsnAsnGln	628
Db	2608	AAGCCTCGTGCAGACATACATCATCTGTCTGTACGGGCGATATGAATATGATCTTATGGAA	2667
Oy	629	LysAspGluTrpLysIleGlySerValSerValaLysThrProThrGlyThrHis	648
Db	2668	GATCCCTCATAGATAGAGTGGCTTCCATCAAAATGACAGACATACAGACACCCATTTCA	2727
Oy	649	ProLeuPheGluValaGlnHisAsnGlnThrSerThrTyrrThrLeuAspMetAlaHisAsn	668
Db	2728	TGCTTACAGCGGGCTCCAAACCCACAGATCATCTGTCTCADAAGTTCAATCCGGGCCAATCC	2787
Oy	669	ArgValaLeuProAspPheThrLeuLysAsnLeu---GlyLeuProPheAsnGlyGluLeu	687
Db	2788	GAGACATATGCCGACTTTCAGCATTAACAATGAAGTGGTGTACTTCTCCGGGAATTC	2847
Oy	688	ValaValaValaPheAlaGlyGlnThrGlnSerSerSerGlySerLeuThrAlaAlaGlnGluThr	707
Db	2848	GAATTAAGTGGCTATTAAGGCTTTCTCGAAACT---TTCTTCAAGGAAAGAAAGAACAC	2904
Oy	708	ValaHisIleLeuGlnGlyIleThrPheValaTyrrGlyLeuProValaValaGlyPro-----	725
Db	2905	ATGAGCTCGCCCAAGGGGAAACCAAGATATGTCTCCGAGCGTACGACGCAACTTTCT	2966
Oy	726	-----IleProAspGlySerTyrrArgAlaThrLeuHisAlaPhe	738
Db	2965	CTGTATACAAATGCCAGAACTCTTCCGATGGCATCTATTAATGTGCATGAGA-----	3018
Oy	739	ValaAsnGlyGlnGlnLeuTyrr-----IleuysGlyLysArgAsnTyrrThr	754
Db	3019	-----GAGCAGGGGATTTGGGATCCGATCGATTTGTTGGGAGTATTAATCATCGT	3069
Oy	755	ValaLysIleValaAsnGlyThrAlaValaGlnAlaIleGluSerSerGlu-----GluIle	772
Db	3070	ATCCGCTCATATTACGAGATCTATCTCTTCCGACATCGCTGTAAAGAGATTTTCTATCTATA	3129
Oy	773	ArgValaPheProAsnProAlaArgAspTyrrAlaGluIleSerAlaProCysIleProGln	792
Db	3130	GTAATTATTCACCATCTGTCTCAGACAGTATCTCATATAGGCATTTCTCCACATATTCG	3189
Oy	793	GluThrSerIleIleLeuPheAspLeuSerGlyLysIleValaMetLysAsnSerLeu-Se	812
Db	3190	GCGACGACACTTGGTGTGTGCATATTCAAAGGCGAATGACAGCTCTCCAGAAAGTCCGA	3249
Oy	812	raIaGlyHisGlyArgMetAspValSerArgLeuProAsnGluValaTyrrIleLeuLysVala	832

Db		3250	ATCGCCGATTGGCTGTCAACGCCTTGAAGGCCACTATATCGTTTGCT	3309
Oy		832	IAspGIYtYrThrThrLysIleasnIIe	841
Db		3310	AGAAGACATGCTAGGAAGCCTCTTAATT	3337
RESULT 4				
STRSPEBJ				
LOCUS				
DEFINITION	STRSPEBJ	1357 bp	DNA	linear BCT 18-MAY-2001
ACCESSION	Streptococcus pyogenes			
VERSION	Streptococcus pyogenes			
KEYWORDS	Bacteria; Filicutes; Lactobacillales; Streptococcaceae;			
SOURCE	Streptococcus pyogenes.			
ORGANISM	Streptococcus pyogenes.			
REFERENCE	Streptococcus pyogenes.			
AUTHORS	Kapur V., Topouzis S., Majesky M.W., Li L.L., Hamrick M.R., Hamill R.J., Patti J.M. and Mosser J.M. A conserved Streptococcus pyogenes extracellular cysteine protease cleaves human fibronectin and degrades vitronectin Microb. Pathog. 15 (5), 327-346 (1993) 94285789			
JOURNAL				
MEDLINE				
PUBMED				
FEATURES				
source	Location/Qualifiers			
	1..1357			
	/organism="Streptococcus pyogenes"			
	/isolate="MGAS 587"			
	/db_xref="taxon:1314"			
gene	161..1357			
	/gene="speB"			
CDS	161..1357			
	/gene="speB"			
	/function="cysteine protease precursor"			
	/note="allele 18"			
	/codon_start=1			
	/transl_table=11			
	/product="pyrogenic exotoxin B"			
	/protein_id="AA27000.1"			
	/db_xref="GI:431656"			
	/translation="MNKKKIGIRLLSLALGGFVLANPVPADONFAHREKANDSAIT PIQSAAIKAKARAEEDIKDKNVLGSELSSNNYYNISTGCVITYSGKRSEILG YSTGSFDANKENIAFMESIYQIEIKNKRLDTTYCIAEIKOPPVKSILDSKGTHY SNQNPYNHLPTVEIKVRGEQSFPVGQAAACVATAQAQIMKHNYPNKLKIDTTYL MNSDPPFNPLFTLEAFISTROYNMNIILPTYSGRSHNVOKMAISELMADGIVSDYD GPSGSAASSFVORALRENFQSYOSVQIINSRDESKODMESDIKELSAQNOPLYOGV GVKYGAVFVIDAGDRNPFHYHMWGMGVSQGFPLDLALNPISALTGGAGGPNFYOSA VVGIKIP"			
BASE COUNT	446 a	262 c	267 g	382 t
ORIGIN				
Alignment Scores:				
Pred. No.:	3.3e-24			1357
Score:	499.00			Matches: 126
Percent Similarity:	50.13%			Conservative: 72
Best Local Similarity:	31.90%			Mismatches: 139
Query Match:	11.24%			Indels: 59
DB:	1		Gaps:	12
US-10-030-330-1 (1-843) x STRSPEBJ (1-1357)				
Oy	LysGIuArGaLaLeuSerLeuAlaArgDeuAlaLeuArgGlnValSerLeuArgMetGly	45		
Db	267 AAGAAGCAAAGATAGCCCTATCATTTATCCAAAATACGCA-GCTATCAAAAGCAGT	325		
Oy	46 GlnThrAlaValSerAPLYSIleSerLLeaspTYrValTYrArGInGlyASPalaGlu	65		
Db	326 GCACGAAGCCGACAAGAT---ATTAAAGCTTGACAAAGTTAACTAGGTGGAGAACTT---	379		
Oy	66 ArgGLylethrSerGIngluGlySerTroAlaTyrrPhetyrValAlaAsnArgGly	85		

Db	380	-----	TCGGGCTCAATGATGTTTCAAAATATTTCTACGTGA	418
QY	86	Asn	snnglucgluyltYrAlaLeuValAlaAlaAsp	105
Db	419	-----	GGATTTGTTATCGTTTCAGAGATTAACCTTCCAGAAATTCGTGAATAC	469
QY	106	Ser	ProIleGlyAlaGlyPheAspMetAspSerMetProAspIleuArgMetTyrLeuGln	125
Db	470	TC	TACACGAGGATCATTTGACGCTAACCGCT---AAGAAACATTTGCTTCTTCATGAA	526
QY	126	Ile	tyr-----AspGlnIuileGlyLeuIleLeuSerGlyLysAla	139
Db	527	AG	TATATGTCGAACAATCAAGAAATATTAAGAAATTAAGCACTACTATATGCTGTGAACGCT	586
QY	140	Gln	LeuAsnnglucgluIleLeuArgThrGluGlyValProAlaGluValHisAlaLeuMet	159
Db	587	GAG	TTAAACACACAGT-----GTTAAATCTCTCCTT	619
QY	160	Asp	AsnnglYHisPheAlaAsnAspProMetArgTyrPaspnglYIlyrProTyrAsnAsn	179
Db	620	GAT	-----TCAAAAGCATTCATTACATCAAGTAAACCTTCAACACTA	664
QY	180	Lys	GluProLeuLeuProAsn-----GlyAsnHisAla	190
Db	665	TTG	ACACCTGTTATGAAAAAGTAAACCAGGTACACATCTTTGTAGTCAACATGCA	724
QY	191	Tyr	ThrGlyCysValAlaThrAlaAlaGlnIleMetArgTyrHisSerTyrProLeu	210
Db	725	GCT	ACAGATGTGTGCTACTGCAACGCTCCAAATTTAAGAAATATCATAAATTACCTAAC	784
QY	211	Gln	IleGlyGluGlySerPheAspTyr-----HisAla	220
Db	785	AA	AGGGTTGAAGACTACACTTACACACTAAGCTCAAAATACCCATATTTCCACACTCT	844
QY	221	Gly	SerLeuValGlyAsnTyrSerGlyThrPheGlyGluMetYrAspTyrIleAsn---	239
Db	845	AAG	AAACTTG-----TTTGCAACATCTCTACTAGACAAATACAACTGGAAACAACTC	895
QY	240	Met	ProGlyAsnProAspPheLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThr	259
Db	896	CT	ACTTCTTATTCACGGAAGAGATCTCAACCTTCAAAAAAATG-----GCATTTACAGA	949
QY	260	Leu	MetArgAspValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThr	279
Db	950	TTG	ATGGCTGATGTGGTATTCAGTACATGGATTTATGGTCCATCATGTAGTGTTCTGCA	1009
QY	280	Tyr	SerValTyrValAlaGlyAlaLeuArgAsnAspPheArgTyrLysArgSerLeuGln	299
Db	1010	GGT	AGCTCTCGTGTTCAAAGAGCCTTGAGAAACCTTGGCTCAACCAACTCTGTTAC	1069
QY	300	Leu	HisValArgAlaLeuTyrThrSerGlnIuItrPHisAspMetIleArgIleGluLeu	319
Db	1070	CAA	ATTCACCTGATGACTTATGGAACCAAAATTTGGGAATCAACAATGACAAAGAAATTA	1129
QY	320	Ala	SerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheVal	339
Db	1130	TC	TCAAAAACCAACAGTATCTACTACCAAGTGTGCGTAAAGTAAAGCGGACATGCTTGT	1189
QY	340	Cys	AspGlyTyrAlaSerAspGlyThrPheHisPheAsnTyrGlyTyrPglyLysValSer	359
Db	1190	ATC	ATGATGCTGCTGACGACGATCACTTACATGTTAACTGGGGTGGGGGAGTCTCT	1249
QY	360	Asn	GlyPheTyrIleuThrLeuLeuSerProThrSerIleuGlyIleGlyGlyGluGly	379
Db	1250	GAG	GGCTCTTCGCTTTCGACGACATTAACCCCTTCAGCTCTGTGTACTGTGGGGCGCGCA	1309
QY	380	Ile	GlyPheThrIleTyrGlnGlnIleIleThrGlyIleGluPro	394
Db	1310	GCG	GCTTCACAGGTTACCAAAAGTGTGTTAGACATCAAACT	1354

LOCUS	STRSPERO	1357 bp	DNA	linear	BCT 18-MAY-2001
DEFINITION	Streptococcus pyogenes pyrogenic exotoxin B (speB) gene, complete cds.				
ACCESSION	L26139				
VERSION	L26139.1	GI:431665			
KEYWORDS					
SOURCE	Streptococcus pyogenes.				
ORGANISM	Streptococcus pyogenes.				
REFERENCE	Bacteria: Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.				
AUTHORS	1 (bases 1 to 1357)				
TITLE	Kapur,V., Topouzis,S., Majesky,M.W., Li,L.L., Hamrick,M.R., Hamill,R.J., Patti,J.M. and Musser,J.M.				
JOURNAL	A conserved Streptococcus pyogenes extracellular cysteine protease				
MEDLINE	cleaves human fibronectin and degrades vitronectin				
PUBMED	Microb. Pathog. 15 (5), 327-346 (1993)				
FEATURES	94385789				
SOURCE	7516997				
gene	Location/Qualifiers				
CDS	1..1357				
	/organism="Streptococcus pyogenes"				
	/isolate="MGAS 650"				
	/db_xref="taxon:1314"				
	161..1357				
	/gene="speB"				
	161..1357				
	/gene="speB"				
	/function="cysteine protease precursor"				
	/note="allele 23"				
	/codon_start=1				
	/transl_table=11				
	/product="pyrogenic exotoxin B"				
	/protein_id="AAZ7005.1"				
	/db_xref="GI:431666"				
	/translation="MNKKKIGRLSLALGCVLANPFAVDQNFARNEKAKDSALIT				
	FIQSAIAKAGARSADIKIDKVNIGELSGSNMYVNIISGGEVYISGDKRPEIIG				
	YSTGSGPANGKENIASPMSYVQIKENKRLDITVGTAEIKQPVYSLIDSGIH				
	NOGNPNILITPVIEKVRKGESPIVQGHATCCVATATQIKRINYPKGLKDTYTT				
	SSNNPHEHKNLEFALSTROYNNMNLPTISGKESNYOKAISLEADVISTDMD				
	GPSSGSAGSSVQKALKEFNGVQNSVHQINSDSKQDWEKQIDKELSQNPVYQV				
	GVKGHAFVIDGADGRNFRYHVMWGMGVSDGFRLDALNPALGTGGAGGNGYQS				
	"VGKIRP"				
BASE COUNT	446 a 264 c 268 g 379 t				
ORIGIN					
Alignment Scores:					
Pred. No.:	3.3e-24	Length:	1357		
Score:	499.00	Matches:	126		
Percent Similarity:	50.38%	Conservative:	73		
Best Local Similarity:	31.90%	Mismatches:	138		
Query Match:	11.24%	Indels:	59		
DB:	1	Gaps:	12		
US-10-030-330-1 (1-843) x STRSPERO (1-1357)					
OY	26	LysGluArgAlaLeuSerLeuAlaArgGluAlaLeuArgInValSerLeuArgMetGly	45		
DB	267	AAGAAGCAAAAGATAGCGCTATCATATTATCCAAAATACAGCA-GCTATCAAGCAGCT	325		
OY	46	GlnThrAlaValSerAspPylsIleSerIleAspTyrValTyrArgGlnGlyAspAlaGlu	65		
DB	326	GCACGAGCGCAGCAAGAT--ATTACGCTTGACCAAGATTAACTTAGGTAGCGAGAACTT--	379		
OY	66	ArgGlyIleThrSerGlnGlnGlnGlnGlnProAlaTyrPheTyrValAlaAsnArgGly	85		
DB	380	-----TCGGCCTCTAATATGATATGTTTACAAATATTCTACTACGGA	418		
OY	86	AsnAsnGlnGlyTyrAlaLeuValAlaAlaAspAspArgIleProThrIleLeuAlaTyr	105		
DB	419	-----GGATTGTTGATCGTTTACAGAGATTAACGTTCTCCAGAAATTCTAGGATAC	469		
OY	106	SerProIleGlyArgPheAspPheCaspSerMetProAsnAsnLeuArgMetTrpLeuGln	125		

```

Db 470 TCTACACGGGATTCATTTGACGCTAACGCT---AAGAAACATTCCTCTTCATCGAA 526
Oy 126 ILETYR-----AspGlnIuilegLYleuIleuSerGlySala 139
Db 527 AGTTATGTCGAACCAATCAAGAAACAAATTAAGACATTAATTCCTGTAACCGCT 586
Oy 140 GlnLeuAnGluGluIleLeuArgThrGluGlyValProAlaGluValHisAlaLeuMet 159
Db 587 GAGATTAAACACCAAGCTT-----GTTAAATCTCTCCCTT 619
Oy 160 AspArgnGlyHisPheHisAlaAsnAspPromeLArgThrPasnGlnGlyTyrProThrPasnAn 179
Db 620 GAT-----TCAAAGGCAATTCATTAACATCAAGTAACCCCTTACCAACCTTA 664
Oy 180 LysGluProLeuLeuProAsn-----GlyAsnHisSala 190
Db 665 TTGACACCTGTTATTTGAAAGTAACCAAGGTGAACATCTTTTGTAGCTACACCTGCA 724
Oy 191 TTTTGTGlyCysValAlaThrAlaAlaAlaGlnIleMetArgTyrHisSerTrpProLeu 210
Db 725 GCTACAGAGATGTGTGCTACTGCAACTGCTCAATATTAATGAAATATCATTAACCTTAC 784
Oy 211 GlnGlyGluGlySerPheAspTyr-----HisAla 220
Db 785 AAGGCTTGAAAGACTACACTTACACACTACCTCAAAATACCATATTTTCAACCATCT 844
Oy 221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetTyrAspTrpLeuAn--- 239
Db 845 AAGAACTTG-----TTGCAAGTATCTCTACTACGCAATACATCAACATGGAACATC 895
Oy 240 MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrLeu 259
Db 896 CTACCTACTTATAGCGGAGAGATCTAACGTTCAAAAAAAGT-----GCGATTTGAGAA 949
Oy 260 LeuMetArgAspValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThr 279
Db 950 TTGATGCGCTGATGTGGTATTTTCAGTACATGATGATTAAGTCCATGCTAGTGTCTGCA 1009
Oy 280 TyrSerValTyrValAlaGlyAlaLeuArgAsnAspPheArgTyrLysArgSerLeuGln 299
Db 1010 GGTAGCTCTCGCTTCAAGACCTTGAAGAAACCTTGGCTACACCAATCTGTCTAC 1069
Oy 300 LeuHisValArgAlaLeuTyrThrSerGlnGluTyrPheHisAspMetIleArgGlyLeu 319
Db 1070 CAATATCAACCGTAGCGACATTAAGCAACAAAGATTGGGAGACCAAAATTGACAAAGATTG 1129
Oy 320 AlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheVal 339
Db 1130 TCTCAAAACCAACGATATACCAAGGTGTGCGTAAGTAAGCGGACATGCTTGTGT 1189
Oy 340 CysAspGlyTyrAlaSerAspGlyThrPheHisPheAsnTrpGlyTyrGlyValSer 359
Db 1190 ATCGATGGTGGTGCAGCGAGCTAACTTCTACATGTTAACTGGGGTGGAGCTCT 1249
Oy 360 AsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGluGly 379
Db 1250 GAGCGCTTCTCGCTTGCACCACTAACCCCTTGCCTGTGCTAGCTGTGGCGGCGCA 1309
Oy 380 IleGlyPheThrIleTyrGlnGluIleIleThrGlyIleGluPro 394
Db 1310 GCGGCTTCAACGCTTACCAAGTGTCTGTGTAAGCATCAAACT 1354

```

```

RESULT 6
STRSPERO 1357 bp DNA linear BCT 18-MAY-2001
LOCUS Streptococcus pyogenes pyrogenic exotoxin B (speB) gene, complete
DEFINITION cds.
ACCESSION L26141
VERSION L26141.1 GI:431669
KEYWORDS
SOURCE Streptococcus pyogenes.
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

```

```

REFERENCE 1 (bases 1 to 1357)
AUTHORS Kapur V., Topouzis S., Majesky M.W., Li L.L., Hamrick M.R.,
Hamil R.J., Patti J.M. and Musser J.M.
TITLE A conserved Streptococcus pyogenes extracellular cysteine protease
cleaves human fibronectin and degrades vitronectin
JOURNAL Microb. Pathog. 15 (5), 327-346 (1993)
MEDLINE 94285789
PUBMED 7516997
FEATURES
source
    1..1357
    /organism="Streptococcus pyogenes"
    /isolate="MGAS 168"
    /db_xref="taxon:1314"
    161..1357
    /gene="speB"
    161..1357
    /gene="speB"
    /function="cysteine protease precursor"
    /note="allele 25"
    /codon_start=1
    /transl_table=11
    /product="pyrogenic exotoxin B"
    /protein_id="AA27007.1"
    /db_xref="GI:431670"
    /translation="MNKKRIGRLSLAAGFVLANPVADQNFARNEKAKDSAIT
    FLOKSAATKAGARSADIKDKVNIAGGELSGNMTIYNISGTFVYISGDRSPETIG
    YSTSSPDYNGKENTASFPESVEDIKERKIDTAYAGAEIKQPVYSLDSKGIHY
    NQGNFNLITPEIKVKEPODSFVQGHATGCVATATQIMKTHNPNGKLDVYTL
    SSNNPYFNHPKMLFAAISTROYNNMNIILPTYSGRSNVOKMISELMDVGLISVMDY
    GSSSGSAGSSRVORALKEFNFGVQSHQIRSDFKOMESOIDKELSONQGVYQVQ
    GVGGHAFVIDADGRNFYHVMGWSVSDGPFRLDALNPALGTGGAGGFGNGVSA
    VYGIKP"
BASE COUNT 448 a 265 c 265 g 379 t
ORIGIN
Alignment Scores:
Pred. No.: 3..3e-24 Length: 1357
Score: 499.00 Matches: 125
Percent Similarity: 50.38 Conservative: 74
Best Local Similarity: 31.65 Mismatches: 138
Query Match: 11.248 Indels: 59
DB: 1 Gaps: 12
US-10-030-330-1 (1-843) x STRSPERO (1-1357)
Oy 26 LysGluArgAlaLeuSerIleuAlaGlyLeuAlaLeuArgGlnValSerLeuArgMetGly 45
Db 267 AAGAAACCAAAAGATACCGCTATCACATTTTACCAAAATCAACA-GCTATCAACGAGGT 325
Oy 46 GlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArgGlnGlyAspAlaGlu 65
Db 326 GCACGAAAGCGCAAGAT---ATTAAAGCTTGCAAAAGTTAACTTAGTGGAGAACT--- 379
Oy 66 ArgGlyIleThrSerGlnGluGlySerProAlaTyrPheTyrValAlaAsnArgGly 85
Db 380 -----TCTGGCTCTAATATATATATATATACAAATATTTCTACTGGA 418
Oy 86 AsnAsnGluGlyTyrAlaLeuValAlaAlaAspAspArgIleProThrIleLeuAlaTyr 105
Db 419 -----GGAATTTGTTATCGTTTCAGAGATAAACGTTCCAGAAATTTCTAGGATAC 469
Oy 106 SerProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMetTrpLeuGln 125
Db 470 TCTACACGGGATTCATTTGACGCTTAACGCT---AAGAAACATTCCTCTTCATCGAA 526
Oy 126 ILETYR-----AspGlnIuilegLYleuIleuSerGlySala 139
Db 527 AGTTATGTCGAACCAATCAAGAAACAAATTAAGACATTAATTCCTGTAACCGCT 586
Oy 140 GlnLeuAnGluGluIleLeuArgThrGluGlyValProAlaGluValHisAlaLeuMet 159
Db 587 GAGATTAAACACCAAGCTT-----GTTAAATCTCTCCCTT 619

```

Oy	160	AspsannglyhisPheAlaAsnAspPrometargtrpransnglytyrProtrPrasasn	179
Db	620	GAT-----TCAAAAGGCATTATTACCAACCAAGACCCTTACCAACTA	664
Oy	180	LysGslupProLeuLeuProasn-----GlyAsnHisIsla	190
Db	665	TTGACACCTGTTATTGAAAAAGTAAACCCAGCTAACACACTTTTGTAGGTACACATGCA	724
Oy	191	TyrThrGlyCysValAlaThrAlaAlaIacInilemetargtyrHisSerTrpProleu	210
Db	725	GCTACAGAGATGTGTGCTACTGCACAAAGCTCAAAATTATGAATATCATAAATACCCATAAC	784
Oy	211	GlnGlyGlnGlySerPheAspTyrr-----HisIsla	220
Db	785	AAAGGGTTGAAAGACTACACTTACACACTAAGCTCAATAAACCATATTTGACCAACTCT	844
Oy	221	GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGlnMetTyrrAspTrpIleasn---	239
Db	845	AAGAAGCTTG-----TTTGACAGCATCTCTACTAGACATATACACTGTGACAAACATC	895
Oy	240	MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrrAlaThr	259
Db	896	CTACCTACTTATTACCGGAGAGAACTCTAACCTCAAAAATG-----GCCATTTTCAGAA	949
Oy	260	LeuMetArgAspValSerAlaSerValSerMetSerPheTyrrGlnAsnGlySerGlyThr	279
Db	950	TTGATGGCGTGGATGTGTGATTTCAGTACATACATGATTAATGGTCATCTAGTGGTCTGCA	1009
Oy	280	TyrSerValTyrrValValGlyAlaLeuArgAsnAsnPheArgTyrrLysArgSerLeuGln	299
Db	1010	GGTAGCTCTCGTGTTCAAAGAGCCTTGAAAGAAACCTTGGCTTACACCAATCTGTTCAC	1066
Oy	300	LeuHisValArgAlaLeuTyrrThrSerGlnGluTrpHisAspMetIleArgGlyLeu	319
Db	1070	CAAAATCAACCCCTAGCTACTTTAGCAACAACAGATGGGAATGCACAATATGCACAAAGAAATTA	1129
Oy	320	AlaSerGlyArgProValTyrrTyrrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheVal	339
Db	1130	TCTCAAAAACCAACAGTATCTACCAAGAGTCTCGGTGATAAGGCGGACATGCCCTTTGTT	1189
Oy	340	CysAspGlyTyrrAlaSerAspGlyThrPheHisPheAsnTrpGlyTyrrGlyValSer	359
Db	1190	ATCGATGGTGTGACGAGCACTTAACCTCTACACATGTAACTGGGGGTGGGGAGTCTCT	1249
Oy	360	AsnGlyPheTyrrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGly	379
Db	1250	GACGCGTCTTCCTCGCTTGAAGCAGCACTAAACCCCTCAGCTCTTGGATACGTGGCGGCGCA	1309
Oy	380	IleGlyPheThrIleTyrrGlnGlnIleIleThrGlyIleGluPro	394
Db	1310	GGCGGCTTCAACGGTTACCAAGTCTGCTGTGTAGCAATCAAAACCTT	1354

[illegible]

FEATURES	source	location/Qualifiers
gene	CDS	1..1357
		/organism="Streptococcus pyogenes"
		/isolate="WGS 165"
		/db_xref="taxon:1314"
		161..1357
		/gene="speB gene"
		161..1357
		/gene="speB gene"
		/function="cysteine protease precursor"
		/note="allele 4"
		/codon_start=1
		/transl_table=1
		/product="pyrogenic exotoxin B"
		/protein_id="AA36986.1"
		/db_xref="GI:431628"
		/translation="MNRKLGIRLLSLLAGGYLANPVRADQNFARNEKEAKDSATTF
		FIQSAIAKGAARSADIEIKLDKVNIGGELSGSNMYVNIISGCFVIYSGKRSEIILG
		YISGSGPDVNGKENIADSFMSYVQIEENRKLDITVTAGTAEIKQPVKSLIDSGIHY
		NOGPNYILTFVIEKRVFGEQSFVGOHAATGCVTATAQIKYHNYPNKGIKDTYTL
		SSNPYFNPKNLPAALSTRQINNNILPTYSGESNVOKAIISELMAVGISVDMDY
		GPSSGASGSVRORALIKENGYNSVHGINSDSKDWSQIDKEISLQNPQVYVGG
		GVKVGCHAFVIDGADGRNPFYHYNMGSGVSDGFRLDLNPSALSTGGGAGGFNGQSA
		VVGIKP"

ORIGIN	BASE COUNT	448 a	264 c	265 g	380 t
Alignment Scores:					
Pred. No.:	3,84e-24	Length:	1357		
Score:	498.00	Matches:	125		
Percent Similarity:	50.38%	Conservative:	74		
Best Local Similarity:	31.65%	Mismatches:	138		
Query Match:	11.22%	Indels:	59		
DB:	1	Gaps:	12		
US-10-030-330-1 (1-843) x STRSPEBAG (1-1357)					
OY	26	LysGLUaRgAlaLeuSerLeuAlaRgLeuAlaLeuArgGlnValSerLeuArgMetGly	45		
Db	267	AAGAAAGCAAAAGATAGCGCTATCCACATTTATCCAAAATACGCA-GCTATCCAAAGCAGGT	325		
OY	46	GlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArgGlnGlyAspAlaGlu	65		
Db	326	GCAGAAAGCAGAGAGAT---ATTAGCTTACCAAGTTAACTTAGGTGGAACTT---	379		
OY	66	ArgGlyIleThrSerGlnGluGlySerProAlaTyrPheTyrValAlaAsnArgGly	85		
Db	380	-----TCTGGCTCAATATGATGTATGTTTCAATATTTCTACGGA	418		
OY	86	AsnAsnGluGlyTyrAlaLeuValAlaAlaAspSparGlyIleProThrIleLeuAlaTyr	105		
Db	419	-----GGATTTGGTATCGTTTCCAGGAATTAACGCTTCTCCAGAAATTTCTAGATAC	469		
OY	106	SerProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMetTyrLeuGln	125		
Db	470	TCTACCAAGCGATCATTTGACGTAAACGCT---AAAGAAACATTTGCTTCCTCATGGAA	526		
OY	126	IleTyr-----AspGlnGluIleGlyLeuIleLeuSerGlyLysAla	139		
Db	527	AGTTATGTGCAACAATCAAGAAACAAACAAAATTAACACTACTATTAGCTGTACCGCT	586		
OY	140	GlnLeuAsnGluGluIleLeuArgThrGlnGlyValProAlaGluValHisAlaLeuMet	159		
Db	587	GAGATTTAAACCAACAGT-----GTTAAATCTCTCTCT	619		
OY	160	AspAsnGlyHisPheAlaAsnAspPromeLeuArgTyrAsnGlnGlyTyrProTyrAsnAsn	179		
Db	620	GAT-----TCAAAAGGATTCATTACATCAAGSTAAACCTTCAACACTA	664		
OY	180	LysGluProLeuLeuProAsn-----GlyAsnHisAla	190		
Db	665	TTTGACACCTGTTATTGAAAAAGTAAACCAAGGTGACATCTTTTGTAGGTCAACATGCA	724		

Oy 191 TyrThrGlyCysValAlaThrAlaAlaGlnIleMetArgTyrHisSerProLeu 210
 Db |||||
 Oy 725 GCTACAGATGTGTCTACTGCACTGCTCAATATGAAATATCAATATATACCTAAC 784
 Db |||||
 Oy 211 GlnGlyGlnGlySerPheAspTyr-----HisAla 220
 Db |||||
 Oy 785 AAGGGCTGAAGACTACACTTACACTTAAGCTCAATATACCATATTCACACATCT 844
 Db |||||
 Oy 221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGlnMetTyrAspTrpIleLeu 239
 Db |||||
 Oy 845 AAGAACTTG-----TTTGAGCTATCTCTACTAGACATATCAACATGGAACACATC 895
 Db |||||
 Oy 240 MetTrpGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThr 259
 Db |||||
 Oy 896 CTACCTACTTATAGCGGAGAGATCTAACCTTCAAAAAATG-----GCGATTTCAAGA 949
 Db |||||
 Oy 260 LeuMetArgAspValSerAlaSerValSerMetSerPheTyrGlnAsnGlySerGlyThr 279
 Db |||||
 Oy 950 TTGATGGCTGATGTGGTATTTCAAGATGATGATGATGATGATGATGATGATGATGAT 1009
 Db |||||
 Oy 280 TyrSerValTyrValValGlyAlaLeuArgAsnAspPheArgTyrLysArgSerLeuGln 1069
 Db |||||
 Oy 1010 GGTAGCTCTGCTGTTCAGAGCCTTGAAGAAAGAACTTGGCTCAACCAATCTGTCTAC 1069
 Db |||||
 Oy 300 LeuHisValArgAlaLeuThrThrSerGlnGlnTyrPheAspMetIleArgGlyLeu 1129
 Db |||||
 Oy 1070 CAATCAACCGTACGACCTTACCAAAAGATGGAGATCAAAATTCACAAAGATTA 1129
 Db |||||
 Oy 320 AlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheVal 1139
 Db |||||
 Oy 1130 TCTCAAAACCAACAGTATCTACCAAGGTGCTCGTAAAGAGCGGACATGCTTTGTT 1189
 Db |||||
 Oy 340 CysAspGlyTyrAlaSerAspGlyThrPheHisPheAsnTrpGlyTyrPheGlyValSer 1190
 Db |||||
 Oy 1190 ATCGATGCTGTGTCAGCGACGTAACCTTCTACCATGTTACTGGGGTGGAGTCT 1249
 Db |||||
 Oy 360 AsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlnGly 1309
 Db |||||
 Oy 1250 GACGGCTCTCCCTCTTGACGCACTAAACCTTCAGCTCTGCTAGTGTGGCGGCCA 1309
 Db |||||
 Oy 380 IleGlyPheThrIleTyrGlnGlnIleIleThrGlyIleGlyPro 394
 Db |||||
 Oy 1310 GCGGCTTCAACGCTTACCAAAAGTGTGTTGATGACATCAAACT 1354
 Db |||||
 RESULT 8
 LOCUS AB051298
 DEFINITION Streptococcus pyogenes spe b gene for pyrogenic exotoxin B,
 complete cds.
 AB051298
 AB051298.1 GI:13591590
 KEYWORDS Streptococcus pyogenes (strain: NZ131) DNA.
 SOURCE Streptococcus pyogenes
 ORGANISM Streptococcus pyogenes
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus
 1 (sites)
 Watanabe, Y. and Ohkuni, H.
 Cloning and expression analysis of Streptococcal pyrogenic exotoxin
 B (Spe-B)/streptococcal cysteine protease (SCP)
 Unpublished
 2 (bases 1 to 1479)
 Watanabe, Y. and Ohkuni, H.
 Direct Submision
 Submitted (16-NOV-2000) Yukino Watanabe, Nippon Medical School
 Institute of Gerontology, Department of Immunology and Infectious
 Diseases, Nakahara-ku Kosugi-cho 1-396, Kawasaki-shi, Kanagawa
 211-8533, Japan (E-mail: wbykn@nmsu.nippon.ac.jp,
 Tel: 81-44-733-1821 (ex. 892), Fax: 81-44-733-1877)
 Location/Qualifiers
 1..1479
 /organism="Streptococcus pyogenes"
 /strain="NZ131"

gene /db_xref="taxon:1314"
 154..1350
 /gene="spe b"
 154..1350
 /gene="spe b"
 /codon_start=1
 /transl_table=11
 /product="pyrogenic exotoxin B"
 /protein_id="BAB40954.1"
 /db_xref="GI:13591591"
 /translation="MKKKLGRLLSLALGFLVLPVPDPAFAPNEKEADDSAIT
 YKSSGDFVNGKESAEIDIKLVNIGELSGSNMXYVNISSGVVYSGKRSPELIG
 NGRFENLTPVIEKVEKGEOSFVEQIKENKRLDTPVAGTAEIKQPVKSLDSKGIH
 SSNPFYFNPKMLFAIISTROYNNNITLPTVSGRESVQKMAISELMADVIGISVDKY
 GPSSSSAGSSRYORALKENPFGVNGSVYQINRSDSKQDMESQIDKELSONPVPYQGV
 GVGGHAFVIGADQGRNFYHNMGMGVSDGFRLDALNPSALGTGAGGFGNGYQSA
 VVGIRP"
 BASE COUNT 490 a 298 c 288 g 403 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4 25e-24 Length: 1479
 Score: 498.00 Matches: 127
 Percent Similarity: 50.50% Conservative: 74
 Best Local Similarity: 31.91% Mismatches: 138
 Query Match: 11.22% Indels: 60
 DB: 1 Gaps: 13
 US-10-030-330-1 (1-843) x AB051298 (1-1479)
 Oy 24 ValThrLysGlyArg-----AlaLeuSerLeuAlaThrAlaLeuArgGlnValSerLeu 42
 Db |||||
 Oy 251 GTACAGAAAAGAGAGCAAGATACCGCTATCATCTTATCCAAAATCAGCA-GCATC 309
 Db |||||
 Oy 43 ArgMetGlyGlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArgGlnGly 62
 Db |||||
 Oy 310 AAAAGCAGTGCAGAGAGCGAGAAAGAT--ATTAACTTCAACAAGTATCTTAGGTGGA 366
 Db |||||
 Oy 63 AspAlaGluArgGlyIleThrSerGlnGlnGlnGlySerProAlaTyrPheThrAla 82
 Db |||||
 Oy 367 GAACCTT-----TCTGGCTCTTAATGATGATGATTTTCAATATT 402
 Db |||||
 Oy 83 AsnArgGlyAsnAsnGlnGlyTyrAlaLeuValAlaAlaAspAspArgLysLeuThrIle 102
 Db |||||
 Oy 403 TCTACTGGA-----GGATTGTTATGCTTTCAGAGATTAACGTTCTCCAGAAATY 453
 Db |||||
 Oy 103 LeuAlaTyrSerProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMet 122
 Db |||||
 Oy 454 CTAGATACACTCTACAGCGGATCATTTGACGTTAACCGT--AAAGAAAACATTCGCTCC 510
 Db |||||
 Oy 123 TrpLeuGlnIleTyr-----AspGlnGlnIleGlyLeuIleLeuSer 136
 Db |||||
 Oy 511 TTTCATGAAAGTATTCGACCAACAAATCAAGAAACAAATAATGACACTTATGCT 570
 Db |||||
 Oy 137 GlyLysAlaGlnLeuAsnGlnGlnIleLeuArgThrGlnGlnGlyValProAlaGluValHis 156
 Db |||||
 Oy 571 GGTACCGCTGAGATTAATAACACAGTT-----GTTAA 603
 Db |||||
 Oy 157 AlaLeuMetAspAsnGlyHisPheAlaAsnAspPheMetArgTrpAsnGlnGlyTyrPro 176
 Db |||||
 Oy 604 TCTCTCTGTGAT-----TCAAAAGCATTCATTATACATCAAGTAACTT 648
 Db |||||
 Oy 177 TrpAsnAsnLysGluProLeuLeuProAsn-----Gly 187
 Db |||||
 Oy 649 TACACACTTATGACACCTGTTATGAAAAGTAAACCGAGTAAACATCTTTTGTAG 708
 Db |||||
 Oy 188 AsnHisAlaTyrThrGlyCysValAlaThrAlaAlaGlnIleMetArgTyrHisSer 207
 Db |||||
 Oy 709 CAACATCAGCTACAGAGATGTGCTACTGCACTGCTCAATATGAAATATATCATAT 768
 Oy 208 TrpProLeuGlnGlyGlnGlySerPheAspTyr----- 218

```

/db_xref="GI:431622"
/translation="MNKKIKGLISLALGCVLANPVPADQNFARNEKEAKDSATIT
FIQSAIKAGARSAEDIKIDRVNIGELSGSNMYVNI STYGGFVIVGSKRSPBELLG
YSTSGSEVDNKENIASFMEYSVQIKENKLDITVAGTAEIKPPVAKSLDSKGIHY
NOGNPDYINLPVIEIKVBPQSGFVGQHAATCGVATPAIQRKNYVNRNGKIKDVTYTL
SSNNPFRHPRKNLEFALAISTRQYNNNNILPTYSGESVQKAKISELMADYISDMYQ
GPGSGAGSSRQVRLKENERGTNOSVQIINSDSKQDWEQIDKEISONQPYQGV
GKVGGAHAFVIDAGGRNFYHVMQMGVSEFFRLDALNPSALGTGGACGPNFYQSA
VVGKIP"
BASE COUNT      448 a      263 c      266 g      380 t
ORIGIN

Alignment Scores:
Pred. No.:      5.21e-24      Length:      1357
Score:          496.00      Matches:      125
Percent Similarity: 50.38%      Conservative: 74
Best Local Similarity: 31.65%      Mismatches: 138
Query Match:    11.18%      Indels:      59
Db:              1      Gaps:      12

US-10-030-330-1 (1-843) x STRSPBAD (1-1357)

OY      26  LysgluAArgAlaLeuSerLeuAlaArgLeuAlaLeuArgInValSerLeuArgMetGly 45
Db      267  AAGAAAGCAAAAGATAGCGCTATTCACATTATCCAAAAATCGCA-GCATATCAAGCAGGT 325
OY      46  GluThrAlaValSerAspLysIleSerIleAspTyrValTyrArgGlnLysPalaGlu 65
Db      326  GCAGAGACGGCAGAAAGAT--ATTAGCTTGACAAAGTTAACTTAAGTGAGAACTT--- 379
OY      66  ArgGlyIleThrSerGlnGluGlySerProAlaTyrPheTyrValAlaAsnArgGly 85
Db      380  -----TCTGGCCCTATATGTATGTATCTTCAATATTCTTCACTGGA 418
OY      86  AsnAsnGluGlyTyrAlaLeuValAlaAlaAspAspArgIleProThrIleLeuAlaTyr 105
Db      419  -----GGATTTGTTATCGTTTCAGGACGATTAACCTTCCAGAAATTCAGATAC 469
OY      106  SerProIleGlyIleArgPheAspMetAspSerMetProAspAsnLeuArgMetTyrLeuGln 125
Db      470  TCTTACCAAGCGGATCATTTTGACGTTAAACGCT--AAAGAAACATTGCTCTTCATGAA 526
OY      126  IleTyr-----AspGlnIleGlyLeuIleLeuSerGlyLysAla 139
Db      527  AGTATATGTCGACAACATCAAGAAACAAAAAAATTAGACACTACTTATGCTGTACCGCT 586
OY      140  GlnLeuAsnGluGluIleLeuAlaArgThrGlnIleValProIleGluValIleAlaLeuMet 159
Db      587  GAGATTTAAACAACACAGT-----GTTAAATCTCTCCTT 619
OY      160  AspAsnGlnHisPheAlaAsnAspProMetArgTyrPasnGlnGlyTyrProTyrPasn 179
Db      620  GAT-----TCAAAAGCGCTATTATTTAACAAGTAACCTTCAACCTTAACCTA 664
OY      180  LysGluProLeuLeuProAsn-----GlyAsnHisAla 190
Db      665  TTGACACCTGTTATTGAAAAAGTAAACACAGGTGACAACTTTTGTGTAGTCACATGCA 724
OY      191  TyrThrGlyCysValAlaThrAlaAlaIleGlnIleMetAlaGlyTyrHisSerTyrProLeu 210
Db      725  GCTCAAGAGATGTGTGCTACTGCAACAGCTCAAAATTATGAATATCATAAATTACCTAAC 784
OY      211  GlnGlyGluGlnYserPheAspTyr-----HisAla 220
Db      785  AAAGGTTGAAGAAGACTTACACTTACACACTAAGCTCAAAATTAACCCATTTTCAACCATCT 844
OY      221  GlySerLeuValGlyAsnTyrSerGlyThrPheGlyGluMetYtyrAspTyrIleAsn-- 239
Db      845  AAGAAGCTG-----TTTGCAAGCTATCTCTACTAGACAAATACAACTGGACACATCATC 895
OY      240  MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThr 259
Db      896  TTACTCTATTATACCGAAGAAATCTTAACCTTAAAAAATG-----GGCATTTTCGAA 949

```

[illegible]

BASE COUNT	448 a	264 c	266 g	379 t
ORIGIN				
Alignment Scores:				
Pred. No.:	5,21e-24	Length:	1,357	
Score:	496.00	Matches:	125	
Percent Similarity:	50.384	Conservative:	74	
Best Local Similarity:	31.654	Mismatches:	138	
Query Match:	11.188	Indels:	59	
DB:	1	Gaps:	12	
US-10-030-330-1 (1-843) x STRSPERAK (1-1357)				
QY	26	LysGLuARGLALeUSeRLeuLALaRGLuALALeUaRGLInValSeRLeUaRgmctg	4	
DB	267	AAGAAACAAAGATGATGCCCTTACATTTATCCAAAATTCAGCA-GCTATCAAAACAGCT	3	
QY	46	GLnThRALaValSeRAspLysIleSeRILeAspTYValTYRArgGLnGLYAspAlaGLu	6	
DB	326	GCACAAACGCGAAGAT--ATTAACTTCGAAGAAATTAACTTAGCGGAACCT--	3	
QY	66	ArgGLYIleThSeRGLnGLuGLYSeRProLaTYRPhetyValAlaAsnArgGLY	8	
DB	380	-----TCGGCTCAATATGATATGTTTAAATATTTCTACAGCA	4	
QY	86	AsnAsnGLuGLYTYRAlaLeuValAlaAlaAspAspArgIleProThRIleLeuAlaTYR	1	
DB	419	-----GGATTTCGTATGCTTTAGAGATTAACCTTCAGAAATTTCTAGCATC	4	
QY	106	SeRProIleGLYArgPheAspMetAspSeRmetProAspAsnLeuArgmetTYRpleuGLn	12	
DB	470	TCATCCAGCGGATCATTTGACGTTAACGCT--AAAGAAACATTCCTTCCTTCATGGA	52	
QY	126	IleTYR-----AspGLnGLuIleGLYLeuIleLeuSeRGLYAsn	13	
DB	527	AGTTAATGCGAACAATCAAAAGAAACAAAATTAAGCACTACTTATGCTGTACCGCT	58	
QY	140	GLnLeuAsnGLuGLuIleLeuArgThrgLGLYValProAlaGLuValHisAlaLeuMet	15	
DB	587	GAAATTAACCAACCAAGTT-----GTAAATCTCTCCTT	61	
QY	160	AspAsnGLYHisPheAlaAsnAspProMetArgTYRAsnGLnGLYTYRProTYRAsnAsn	17	
DB	620	GAT-----TCAAAAGCATTCATTAACATCAAGTAACCTTCACCTA	66	
QY	180	LysGLuProLeuLeuProAsn-----GLYAsnHisAla	19	
DB	665	TTCACACCTGTTATTAAGAAAGTAACCAAGTGAACAAATCTTTGATGCTACACATGCA	72	
QY	191	TYRThrgLycValAlaThnRALaAlaAlaGLnIleMetArgTYRHisSeRTTYRleu	210	
DB	725	GCTAAGAGATGTGTGCTACTGCACTGCTCAAAATTAATGAATATCAATTAATACCTAC	784	
QY	211	GLnGLuGLuGLYSeRPhAspTYR-----HisAla	220	
DB	785	AAAGGTTGAAGACTACACTTACACACTACCTCAATTAACCATATTTCAACCACTCT	844	
QY	221	GLYSeRLeuValGLYAsnTYRSeRGLYThrPheGLYGLuMetTYRAspTYRPIleAsn--	239	
DB	845	AAGAAGCTG-----TTTGACGCTATCTCTACTAGACATTAACCACTGGAACCAATC	895	
QY	240	MetProGLYAsnProAspLeuAspAsnLeuThrgLInSeRGLnValAspAlaTYRAlaThr	259	
DB	896	CTACCTACTTAAACCGGAAGAAACAACTTAACCTTCAAAAATG-----CGAATTCAGCA	949	
QY	260	LeuMetArgAspValSerAlaSerValSerMetSeRPhetyTYRGLuAsnGLYSeRGLYThr	279	
DB	950	TTGATGGCTGATGTTGTGATTTCTAGACATGGATATGTGTCACATCTGTGCTTCGCA	1009	
QY	280	TYRSeRValTYRValGLYAlaLeuArgAsnAsnPheArgTYRGLYAsnLeuGLn	299	
DB	1010	GGTAGCTCTCGTTCCAAAGAGCCTTAAAGAAACATTGGGCTACACCAATCTGTTAC	1069	

QY 300 LeuHISValArgAlaLeuTyrThrSerGlnGluTrpHisAspMetIleArgGlyGluLeu 319
DB 1070 CAAATCAACCGTAGACTTACGAAACAGATGCGAAGCAATTAAGCAAAATTA 1129
QY 320 AlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheVal 339
DB 1130 TCTCAAAACCAACGATGACTACCAAGTGTGGTAAAGTAGGAGCATCCCTTGT 1189
QY 340 CysAspGlyTyrAlaSerAspGlyThrPheHisPheAsnTrpGlyTyrGlyValSer 359
DB 1190 ATCGATGGTGTGACGACGACGATTAACCTGTTAACTGAGGTGGGTGGAGTCTCT 1249
QY 360 AsnGlyPheTyrTyrLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGlu 379
DB 1250 GACGGCTCTTCGCTGACGACGACGATTAACCTGAGCTGTGACTGAGTGGCGCA 1309
QY 380 IleGlyPheThrIleTyrGlnGluIleIleThrGlyIleGluPro 394
DB 1310 GCGCGCTTCAACGCTTACCAAAAGTGTGTGAGCATCAACCT 1354

LOCUS STPSPEBC 1357 bp DNA linear BCT 18-MAY-2001
DEFINITION Streptococcus pyogenes pyrogenic exotoxin B (speB) gene, complete cds.
ACCESSION L26127
VERSION L26127.1 GI:431641
KEYWORDS
SOURCE Streptococcus pyogenes.
ORGANISM Streptococcus pyogenes.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE 1 (bases 1 to 1357)
AUTHORS Kapur V., Topouzis S., Majesky M.W., Li L.L., Hamrick M.R., Hamill R.J., Patti J.M. and Musser J.M.
TITLE A conserved Streptococcus pyogenes extracellular cysteine protease cleaves human fibronectin and degrades vitronectin
JOURNAL Microb. Pathog. 15 (5), 327-346 (1993)
MEDLINE 94285789
PUBMED 7516997

FEATURES
source
1..1357 Location/Qualifiers
/organism="Streptococcus pyogenes"
/isolate="MGAS 289"
/db_xref="taxon:1314"
161..1357
/gene="speB"
161..1357
/gene="speB"
/gene="speB"
/function="cysteine protease precursor"
/note="allele 12"
/codon_start=1
/transl_table=1
/product="pyrogenic exotoxin B"
/protein_id="AAA26993.1"
/db_xref="GI:431642"

BASE COUNT 447 a 263 c 266 g 381 t
ORIGIN

Alignment Scores:
Pred. No.: 5,21e-24 Length: 1357
Score: 496.00 Matches: 125
Percent Similarity: 50.13% Conservative: 73
Best Local Similarity: 31.65% Mismatches: 139
Query Match: 11.18% Indels: 59

DB: 1 Gaps: 12
US-10-030-330-1 (1-843) x STPSPEBC (1-1357)
QY 26 LysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgGlnValSerLeuArgMetGly 45
DB 267 AAGAAGCAAAAGATAGCCCTATCATATTATCCAAAATACACA-GCATCAAAAGCAGGT 325
QY 46 GlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArgGlnGlyAspAlaGlu 65
DB 326 GCACAGAGCGAGAAAGAT--ATTAGCTTGTGACAAAGTTAACTTAAGTGGAGAACT-- 379
QY 66 ArgGlyIleThrSerGlnGluGluSerProAlaTyrPheTyrValAlaAsnArgGly 85
DB 380 -----TCTGGCTTAATATGATATGTTTACAAATTTCTACTGCA 418
QY 86 AsnAsnGluGlyTyrAlaLeuValAlaAlaAspArgIleProThrIleLeuAlaTyr 105
DB 419 -----GGATTGTATTCGTTTCAGAGATTAACGTTCCACAGAAATTCAGAGTAC 469
QY 106 SerProIleGlyArgPheAspMetLysSerMetProAspAsnLeuArgMetTrpLeuGln 125
DB 470 TCTACACGCGATCATTTGACGCTAACGCT--AAAGAAAGATTGCTTCCATCAGTA 526
QY 126 IleTyr-----AspGlnGluIleGlyLeuIleLeuSerGlyLysAla 139
DB 527 AGTTATGTGCAACCAATCAAGAAATTAATAATTAAGACATCTTATCTGTCGCT 586
QY 140 GlnLeuAsnGluGluIleLeuArgThrGlnGlyValProAlaGluValHisAlaLeuMet 159
DB 587 GAGATTAAACCAACGCT-----GTTAATCTGCTCCT 619
QY 160 AspAsnGlyHisPheAlaAsnAspPheMetArgTyrPheGlnGlyTyrProThrPheAsn 179
DB 620 GAT-----TCAAAAGCATTCATTAACCAAGTAACCAAGTAACCAAGTAACCA 664
QY 180 LysGluProLeuLeuProAsn-----GlyAsnHisAla 190
DB 665 TTGACACCGTATTGTAATAAATAAACCAGGTGAACCAATCTTGTAGCTCAACATGA 724
QY 191 TyrThrGlyCysValAlaThrAlaAlaAlaGlnIleMetArgTyrHisSerTrpLeu 210
DB 725 GCTACAGATGATGTTGCTCTCTCAACTGCTCAAAATTAATGAAATCATTAATCACTAAC 784
QY 211 GlnGlyGluGlySerPheAspTyr-----HisAla 220
DB 785 AAGGAGTGAAGACTCACTTACACACTAACCTCAATATCAACCATTTTCAACATCTCT 844
QY 221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetTyrAspTrpIleAsn--- 239
DB 845 AAGAACTTG-----TTTCACCTATCTCTACTAGACATTAACACTGGAACCAATC 895
QY 240 MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThr 259
DB 896 CTACCTACTTATAGCGGAAGAAATCAATCAAGTTCAAAAAATG-----CCGATTTCAGAA 949
QY 260 LeuMetAlaAspValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThr 279
DB 950 TTGATGGCTGATGTTGTTATTCATGATGATGATTAATGATGATGATGATGATGATGAT 1009
QY 280 TyrSerValTyrValValGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGln 299
DB 1010 GGTACTCTCGCTTCAAAAGACCTTGAAGAAACTTGGGTACACCAATCTGTCTAC 1069
QY 300 LeuHisValArgAlaLeuTyrThrSerGlnGluTrpHisAspMetIleArgGlyGluLeu 319
DB 1070 CAAATCAACCGTAGACTTACGAAACAGATGCGAAGCAATTAAGCAAAATTA 1129
QY 320 AlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheVal 339
DB 1130 TCTCAAAACCAACGATGACTACCAAGTGTGGTAAAGTAGGAGCATCCCTTGT 1189
QY 340 CysAspGlyTyrAlaSerAspGlyThrPheHisPheAsnTrpGlyTyrGlyValSer 359

[illegible]

326 GCACCAACCGCAGAAAT---ATTAAAGCTTCACAAAGTTAACTTAGGTGAGAACTT--- 379
 66 ATGGIYLIEIHRSERGINGLUGLYSERPROLAIATYRPHETTYVALAIAASNAARGLY 85
 380 -----TCGGCTCTAAATATGATATGTTTACAAATATTTCTACTGGA 418
 86 AASNAAGLUGLYTYRATALEUVALAIAAIAASPARFGLIEPRHRIILEUAIATYR 105
 419 -----GGATTTCGTTATCGTTTACAGAGATAACGTTCTCCAGAAATTTAGCATAC 469
 106 SERPROLIEIATYRPHASPMETASERMETPROASPSLEUARYMETRPLEUGIN 125
 470 TCTACACAGCGGATCATTTGACGGTAAACGTT---AAGAAATAATTCGTTCTCATGGAA 526
 126 ILETYR-----ASPGINUIILEGLYLEUILEUSERGLYLSALA 139
 527 AGTTATGTCGAACAAATCAAGAAACAAATAATTAAGACTACTTATGCTGTGACCGCT 586
 140 GINLEASNGLUGLIULEUARYGTHICLUGLYVALPROLAGLUAVALHISALEUOMET 159
 587 GAGATTAACAACACCGATT-----GTTAAATCTCTCTT 619
 160 ASPASNGLYHISPHLEAIAASNAAPROMELATYRTPASNGINGLYTYRPROTPASASN 179
 620 GAT-----TCMAAGGATTCATTACACACCAAGSTAAACCTTCAACCTA 664
 180 LYSGLUPROLEUENUPROASN-----GLYASNHISALA 190
 665 TTGCACACCTGTATTGTAAGAAAGTAACCGGTGAACAACTTTTGTAGCTCAACATGCA 724
 191 TYRTHGICLYCYVALIATPHRIALAIAGLIIILEMELATYRHISERTRPROLEU 210
 725 GCTACACAGAGTCTCTCTACTGCACTGCACTCAATATGTAATATCATATTAACCTAAC 784
 211 GINGLUGLUGLYSERPHEASPTYR-----HISALA 220
 785 AAGAGTTTAAAGACTACACTTACACACTAACCTCAATTAACCAATATTCACACACTCT 844
 221 GLYSERLEUVALIGLYASNPTRSERGLYHRHIEGLYIUMETTYRASPTRPLEASN--- 239
 845 AAGAACTTG-----TTTGACGCTATCTCTCTACACAACTTCAACCTGGACACCAATC 895
 240 METPROGLYASNPROMASPLEUASPSINLEUTHGINSERGINVALAIPALATYRALATHR 259
 896 CTACCTACTTTACCGGACAGAAAGAACTTCAACGTTCAAAAATAG-----CGGATTTCAGAA 949
 260 LEUETHIATYRSPVALISERLASERVALSERMETSERPHETRYLUSNGLYSERGLYTHR 279
 950 TTTATGGCTATGTTGGTATTATTCAGTACATGATATGGTCAATCATGTGGTCTGCA 1009
 280 TYRSEVALTYRVALGIALYALIAUASNAASPHAEATYRGLYSATRSERLEUGIN 299
 1010 GGTATGCTCTGCTTCAAGAGCGCTTGAAAGAAACTTGGCTCAACCAACTGCTTCAAC 1069
 300 LEUHIISVALATVALALEUITYTHRSERGINGIUTRPHIASPMETIIIEARGLYGLEUEN 319
 1070 CAATATCAACCCGATAGCACTTTAGCAACAAAGATTTGGGAATCAACAATTCAGCAAAAGAAATTA 1129
 320 ALASERGLYATRPROVALTYRGLYLAAGLYASNSANGINSERILEGLYHISALAPHEVAL 339
 1130 TCTCAAAACCAACAGTATCTACCAAGGCTGCGGTAAAGTACGGGACATATGCTTTGTT 1189
 340 CYSASPTGLYTYRALASERASPTGLYTHRPHENIASPHEASNPTRGLYTRPGLYLYVALSER 359
 1190 ATCATAGGTGCTGACGACGACSTAACTTTTACACATGTTAACTGGGGTGGGGTGAAGCTCT 1249
 360 ASNGIYRPHETRYLSLEUTHIRLEUUSERPROTHRSETRLEUGIYILEGLYGLYGLY 379
 1250 GACGCGCTTCTTCGCTTACGACGACCAATTAACCCCTTTCGCTTGTGCTGCGGCGCA 1309
 380 ILEGYPHERHTHLETYRGLINGIUILEIETHGIIILEGLUPRO 394

KEYWORDS
SOURCE Streptococcus pyogenes.
ORGANISM Streptococcus pyogenes
 Bacteria: Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
REFERENCE 1 (bases 1 to 1357)
 Kapur, V., Topouzis, S., Majesky, M.W., Li, L.L., Hamrick, M.R.,
 Hamill, R.J., Patti, J.M. and Musser, J.M.
TITLE A conserved Streptococcus pyogenes extracellular cysteine protease
 cleaves human fibronectin and degrades vitronectin
JOURNAL Microb. Pathog. 15 (5), 327-346 (1993)
MEDLINE 94285789
PUBMED 7516997
FEATURES
 source
 1. 1357
 /organism="Streptococcus pyogenes"
 /isolate="MGAS 800"
 /db_xref="taxon:1314"
 161. 1357
 /gene="speB"
 161. 1357
 /gene="speB"
 /codon_start=1
 /transl_table=11
 /product="pyrogenic exotoxin B"
 /protein_id="AA027002.1"
 /db_xref="GI:431660"
 /translation="MNKKRLGIRLILALGFEVLANPVPADONFARNREKAKDSAT
 /FOKSAIRKAGRSADRIKIDKVNIGELSGSNMYVYNISGSPVIVSGDKRSPILIG
 YSTSSPYNCKENIASFRESVEIOIKERKIDITTYAGTAETKQYVVSILDSKIH
 NOGNPYNLTPYIEKPEQSVQOHATGCVATATQIMKYHNPYKGLDYITL
 SSNNPYFNHFKMLFAIISTROYNNNNILPTYSGRSNVQKAIISMDVGLISVMDY
 GSSSGASSRVOALKEFNQGVQVHOINRSDFSKQDEAOIDKELSONCPYVQGV
 GVGGAHAFVIDADGRNFEYHVMGNGVSDGFRLDALNPALGSGGAGNGYQSA
 YVGRP"
 BASE COUNT 446 a 265 c 266 g 380 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 5.21e-24 Length: 1357
 Score: 496.00 Matches: 125
 Percent Similarity: 50.388 Conservative: 74
 Best Local Similarity: 31.658 Mismatches: 138
 Query Match: 11.188 Indels: 59
 DB: 1 Gaps: 12
 US-10-030-330-1 (1-843) x STRSPEBL (1-1357)
 26 LysGLUArgAlaLeuSerLeuAlaValSerLeuArgMetGly 45
 ||||| : : : : :
 267 AAGAAAGCAAAAGATAGCGCTATCACAATTATCCAAAATCAAGCA-CTATATCAAAAGCAGGT 325
 : : : : :
 46 GlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArgGlnGlyAspAlaGlu 65
 : : : : :
 326 GCACGACGCGCAGAGAT---ATTAAAGCTTGACAAAGTTAACTTAGTGAGGAGAACTT--- 379
 : : : : :
 66 ArgGlyIleThrSerGlnGlnGlnGlySerProAlaTyrPheTyrValAlaAsnArgGly 85
 ||||| : : : : :
 380 -----TCTGGCTCTAATATGATGATGTTACAAATATTCTTCTACTGGA 418
 : : : : :
 86 AsnAsnGlnGlyTyrAlaLeuValAlaAlaAspAspArgIleProThrIleLeuAlaTyr 105
 ||||| : : : : :
 419 -----GGATTGTGATATCGTTTCAGAGATAACGTTCTCCAGAAATTTCTAGGATAC 469
 : : : : :
 106 SerProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMetTrpLeuGln 125
 ||||| : : : : :
 470 TCTACCACGCGATCATTTGACGTTAAGCT---AAAGAAATCATTTGCTTCTCATAGGA 526
 : : : : :
 126 IleTyr-----AspGlnGlnIleGlyLeuIleLeuSerGlyLysAla 139
 ||||| : : : : :
 527 AGTTATGTGCAACAATCAAGAAACAAATAATAGACACACTTATAGCTGTGACCGGT 586
 : : : : :
 140 GlnLeuAsnGlnGluIleLeuArgThrGlnGlyValProAlaGluValHisAlaLeuMet 159

DB : : : : :
 587 GAGATTAAACAACAGCTT-----GTTAAATCTCTCCTT 619
 ||||| : : : : :
 160 AspAsnGlnGlyHisPheAlaAsnAspProMetArgTyrPasnGlnGlyTyrProThrAsn 179
 ||||| : : : : :
 620 GAT-----TCAAAAGCATTCATTACAAACCAAGTACACCTTACACCTTA 664
 : : : : :
 180 LysGluProLeuLeuProAsn-----GlyAsnHisAla 190
 ||||| : : : : :
 665 TTGACACCGCTGATTAATGAATAAAGTGAACAATCTTTGTAGTCAACATGCA 724
 : : : : :
 191 TyrThrGlyCysValAlaThrAlaAlaGlnIleMetArgTyrHisSerTrpLeu 210
 ||||| : : : : :
 725 GCTACAGAGATGTTGCTTCTACTCCAACTGCTCAAAATTAATCAATATACCTTAAT 784
 : : : : :
 211 GlnGlnGlnGlySerPheAspTyr-----HisAla 220
 ||||| : : : : :
 785 AAAGGTTGAAGACTACACTTACACACTAACCTCAATATACCATATTTCAACATCTCT 844
 : : : : :
 221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGlnMetTyrAspTrpIleAsn--- 239
 ||||| : : : : :
 845 AAGAACTTG-----TTTGCACGCTATCTCTACTAGACATTAACACGAGAACATC 895
 : : : : :
 240 MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThr 259
 ||||| : : : : :
 896 TTACCTACTTATAGCGGAGAAAGAAATCTAACCTTCAAAAATG-----GCGATTTTCAGAA 949
 : : : : :
 260 LeuMetArgAspValSerAlaSerValSerMetSerPheTyrGlnAsnGlySerGlyThr 279
 ||||| : : : : :
 950 TTGATGCGCTGATGCTGATTTTCAAGTACATGATATATGCTCAATGATGCTTCTGCA 1009
 : : : : :
 280 TyrSerValTyrValAlaGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGln 299
 ||||| : : : : :
 1010 GGTAGCTCTCGGTTCAAAAGACCTTGAAGAAACCTTGGCTCAACCAATCTGTCTAC 1069
 : : : : :
 300 LeuHisValArgAlaLeuTyrThrSerGlnGlnTyrPheHisAspMetIleArgGlyLeu 319
 ||||| : : : : :
 1070 CAATATCAACCGTAGCGACTTATGCAACAAAGATTTGGAAGCAACAAATGACAAATA 1129
 : : : : :
 320 AlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheVal 339
 ||||| : : : : :
 1130 TCTCAAAACCAACGATATACCAAGGTGCGGTAAAGTGGCGGCATCTTGTGT 1189
 : : : : :
 340 CysAspGlyTyrAlaSerAspGlyThrPheHisPheAsnTrpGlyTyrGlyValSer 359
 ||||| : : : : :
 1180 ATCGATGCGTGCAGCAGCAGTAACTTCACTGATGAGTGGGTGGGTGAGTCTCT 1249
 : : : : :
 360 AsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyLeuGly 379
 ||||| : : : : :
 1250 GACGCGCTTCTTCGCTTGCAGCAGCACTAAACCTTACGCTTGTGACTGGTGGCGGCA 1309
 : : : : :
 380 IleGlyPheThrIleTyrGlnGlnIleIleThrGlyIleGluPro 394
 ||||| : : : : :
 1310 GCGCGCTTCAACGCTTACCAAGTGTCTGTGCGCATCAAACT 1354
 : : : : :
 RESULT 15
 STRSPEBL 1357 bp DNA linear BCT 18-MAY-2001
 LOCUS Streptococcus pyogenes pyrogenic exotoxin B (speB) gene, complete
 DEFINITION cds
 ACCESSION L26137
 VERSION L26137.1 GI:431661
 KEYWORDS
 ORGANISM Streptococcus pyogenes.
 SOURCE Streptococcus pyogenes
 Bacteria: Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
REFERENCE 1 (bases 1 to 1357)
 Kapur, V., Topouzis, S., Majesky, M.W., Li, L.L., Hamrick, M.R.,
 Hamill, R.J., Patti, J.M. and Musser, J.M.
TITLE A conserved Streptococcus pyogenes extracellular cysteine protease
 cleaves human fibronectin and degrades vitronectin
JOURNAL Microb. Pathog. 15 (5), 327-346 (1993)

ORIGIN

Alignment Scores:

Pred. No.: 6,06e-24 Length: 1357
 Score: 495.00 Matches: 125
 Percent Similarity: 50.13% Conservative: 73
 Best Local Similarity: 31.65% Mismatches: 139
 Query Match: 11.15% Indels: 59
 DB: 1 Gaps: 12

US-10-030-330-1 (1-843) x STREPEBAJ (1-1357)

QY 26 LysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgGlnValSerLeuArgMetGly 45
 DB 267 AAGAGCAAAAGATAGGCGTATCATTTATCCAAATATGCGA-GCTATCAAGACAGT 325
 QY 46 GlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArgGlnGlyAspAlaGlu 65
 DB 326 GCACGAAGCCGACGAAGAT---ATTAGCTTGACAAAGTAACTTAGGTGAGAACTT--- 379
 QY 66 ArgGlyIleThrSerGlnGluGluSerProAlaTyrPheTyrValAlaAsnArgGly 85
 DB 380 -----TCTGCGCTAATATGATGTTTACAAATTTCTCTGCA 418
 QY 86 AsnAsnGluGlyTyrAlaLeuValAlaAlaAspAspArgIleProThrIleLeuAlaTyr 105
 DB 419 -----GGATTTGTTATCGTTTCAGAGATAAACGTTCCACAAATTCAGATAC 469
 QY 106 SerProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMetIlePheGln 125
 DB 470 TCTACCAAGCGATCATTTGACGTACGCT---AAGAAACAACTTCCTTCATGAGAA 526
 QY 126 IleTyr-----AspGlnIleGlyLeuIleLeuSerGlyLysAla 139
 DB 527 AGTTATGTGCAACAAATCAAGAAACAAATTAAGACATTAATGCTGTGTAACCT 586
 QY 140 GlnLeuAsnGluGluIleLeuArgThrGluGlyValProAlaGlnValHisAlaLeuMet 159
 DB 587 GAGATTAAACCAACGAT-----GTTAAATCTCTCCTT 619
 QY 160 AspAsnGlyHisPheAlaAsnAspProMetArgTyrPheAsnGlnGlyTyrProTyrAsnAsn 179
 DB 620 GAT-----TCAAAAGCATTTCTATCAACCAAGTAACCTTACCACTA 664
 QY 180 LysGluProLeuLeuProAsn-----GlyAsnHisAla 190
 DB 665 TTGACACCGCTGTTATGAAAGATAAACCCAGGTGCAACATCTTTGTAGTCAACATGCA 724
 QY 191 TyrThrGlyCysValAlaIleThrAlaAlaGlnIleMetArgTyrHisSerTyrProLeu 210
 DB 725 GCTACAGGATGTGTGCTACTGCAACCTGCTCAAAATTATGAAATATCATATACCTTAC 784
 QY 211 GlnGluGluGluSerPheAspTyr-----HisAla 220
 DB 785 AAGGGTGAAGAGACTACACTTACACACTGCAACATATCCATATTTCAACACATCT 844
 QY 221 GlySerLeuValGlyAsnTyrSerGlyThrPheGlyGluMetTyrAspTyrIleAsn--- 239
 DB 845 AAGAACTT-----TTGCAAGCTATGCTCTACTGCAAAATACCACTGCAACATC 895
 QY 240 MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThr 259
 DB 896 TTACCTCTATTATACCGAAGAGATCTACGCTTCAAAATAATG-----GCGATTTCAGAA 949
 QY 260 LeuMetArgAspValSerAlaSerValSerMetSerPheTyrIleLysGlnGlyTyr 279
 DB 950 TTGATGCGCTGATGTTGATATTCAGTACATGATGATATGATGCTGATGCTTTCGA 1009
 QY 280 TyrSerValTyrValGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGln 299
 DB 1010 GGTAGCTCTGCTGTTCAAAAGAGCTTGAAGAAACTTTGGCTACACCAACATCTGTTCAC 1069
 QY 300 LeuHisValArgAlaLeuTyrTyrThrSerGlnGluTyrHisAspMetIleArgGlyLeu 319

DB 1070 CAATACACCGTAGACCTTTAGCAACAGATTTGGAGACACAAATTTGACAAACATTA 1129
 QY 320 AlaSerGlyArgProValTyrTyrAlaGlnAsnAnGlnSerIleGlyHisAlaPheVal 339
 DB 1130 TCTCAAAACCAACCAAGTATACACAGAGTGTTCGTTAAAGTAGCGGACATGCTTGTGT 1189
 QY 340 CysAspGlyTyrAlaSerAspLysThrPheHisPheAsnTyrPglTyrGlyValSer 359
 DB 1190 ATGATGAGGTGCTGACGAGACGTAACTTACATGTTAACTTGAGGTGGGTGAGATCTCT 1249
 QY 360 AsnGlyPheTyrLysLeuThrLeuSerProThrSerLeuGlyIleGlyGlyGly 379
 DB 1250 GACGCGCTCTTCCGCTTGTGACGCACTAAACCTTCAAGCTTGTGTGACTGTGTGCGGCA 1309
 QY 380 IleGlyPheThrIleTyrGlnGluIleIleThrGlyIleGluPro 394
 DB 1310 GCGGCTTCAACGCTTACCAAGTGTGTTGATGATGATCAACAT 1354
 RESULT 19
 STREPEBAJ 1357 bp DNA linear BCT 31-JAN-2002
 LOCUS Streptococcus pyogenes pyrogenic exotoxin B (speB) gene, complete
 DEFINITION cds
 ACCESSION L26160.1 GI:431633
 VERSION L26160.1
 KEYWORDS
 SOURCE
 ORGANISM
 Streptococcus pyogenes.
 Streptococcus pyogenes
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 1 (bases 1 to 1357)
 Kapur, V., Topouzis, S., Majesky, M.W., Li, L.L., Hamrick, M.R.,
 Hamill, R.J., Patti, J.M. and Musser, J.M.
 A conserved Streptococcus pyogenes extracellular cysteine protease
 cleaves human fibronectin and degrades vitronectin
 Microb. Pathog. 15 (5), 327-346 (1993)
 PUBMED 94285789
 7516997
 FEATURES
 source
 Location/Qualifiers
 1..1357
 /organism="Streptococcus pyogenes"
 /isolate="MGAS 429"
 /db_xref="taxon:1314"
 /note="identical sequence found in MGAS 1719"
 161..1357
 /gene="speB"
 161..1357
 /gene="speB"
 /function="cysteine protease precursor"
 /note="allele 7"
 /codon_start=1
 /transl_table=11
 /product="pyrogenic exotoxin B"
 /protein_id="AA26989.1"
 /db_xref="GI:431634"
 /translation="MNKKKLGIRLSILGFIYANVPADQNFARNEKAKSAIT
 FIOKSAIKAGRSAMEDIKLDKVNIGSLSSNNYVNIISGTVIVSGDRSEILG
 YSTSGSPVNGKRENKASFMESYVEDIKENKLDVTVAGTAIKQPVKSLDSKGIH
 NOGNPNYNTLITVEYKRGEDSFVGOHATCGVATAQAIKRYHNPNGIKIDYTYL
 SSNRTFNPKNLPAISTROYNNNLTPTYSGRSHSVOKAISELMADYGISVDMY
 GPSSGASSSRVOKALENFCYNOISVDINGDFSKDWEKQIDKELSONOPVYOGV
 GKVGHAFVIDGADGRNRYHYNMGMGVSDGFPLDLNLSALCTGGGAGGFNGYDQA
 VVGKIP"
 BASE COUNT 447 a 263 c 267 g 380 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 6,06e-24 Length: 1357
 Score: 495.00 Matches: 125
 Percent Similarity: 50.13% Conservative: 73
 Best Local Similarity: 31.65% Mismatches: 139
 Query Match: 11.15% Indels: 59

[illegible]

Db	1190	ATCGATGGTGGCTGACGACGACGTAACTTCTACCATGTATTA	CTGGGGTGGGGTGGAGATTCT	1249
Oy	360	AsnGLpHeTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyValGluGly		379
Db	1250	GACGGCTTCTCCGCTCTGACGACGACGTAAACCTTCAGCTCTTGGTACTACG	GTGGCGGCCA	1309
Oy	380	IleGlyPheThrIleTyrGlnGluIleIleThrGlyIleGluPro		394
Db	1310	GCGGGCTTCAACGGGTACCAAAAGTGTGTGTAGACATCAAACT		1354
RESULT 20				
STRSPEBH		1357 bp	DNA	linear
LOCUS				BCT 18-MAY-2001
DEFINITION		Streptococcus pyogenes pyrogenic exotoxin B (speB) gene, complete cds.		
ACCESSION		L261332		
VERSION		L261332.1		
KEYWORDS		GI:431651		
SOURCE				
ORGANISM		Streptococcus pyogenes.		
REFERENCE		Streptococcus pyogenes.		
AUTHORS		Bacteriella: Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.		
TITLE		1 (bases 1 to 1357)		
JOURNAL		Kapur,V., Topouzis,S., Majesky,M.W., Li,L.L., Hamrick,M.R., Hamill,R.J., Patti,J.M. and Müsser,J.M.		
MEDLINE		A conserved Streptococcus pyogenes extracellular cysteine protease cleaves human fibronectin and degrades vitronectin		
PUBMED		Microb. pathog. 15 (5), 327-346 (1993)		
FEATURES		94285789		
source		7516997		
gene		Location/Qualifiers		
CDS		1..1357		
		/organism="Streptococcus pyogenes"		
		/isolate="MGAS 432"		
		/db_xref="taxon:1314"		
		161..1357		
		/gene="speB"		
		161..1357		
		/gene="speB"		
		/function="cysteine protease precursor"		
		/note="allele 16"		
		/codon_start=1		
		/transl_table=1		
		/product="pyrogenic exotoxin B"		
		/protein_id="AA26998.1"		
		/db_xref="GI:431652"		
		/translation="MNKKKGIRLSLAAGFYLANPYFADQNFARNEKRSATTT		
		FTQSAIRKAGARSAEDIKIDKVLNGLSGSNMYVINISGTFVIVSGDKRSEILIG		
		YSTSGFDANGKENIASFMEYSYEQIENKKLIDTVAGTAEIKQPVVKSILDSKGIHY		
		NOGPNLITPVIEKVRQEQSFVGOHAATCAVTAQAIKRYNPNKGVADTYTLTL		
		SSNSGPNHNPENLFAALSTROYNNNNLIPYSGESNVOKKAISELMAQIDISMDQV		
		GSSGSGSSSRVQALSTREQYINOSYHOINRSDSKDMEQIDKLSQNDPYVQGV		
		GKVGCAHFAVIDGADGRNFRHYHNMGMGVSDGFRLDLNPSALCTGGAGGFGNGQSA		
		VVGIKP"		
BASE COUNT	448 a	265 c	267 g	377 t
ORIGIN				
Alignment Scores:				
Pred. No.:	6.06e-24	Length:	1357	
Score:	495.00	Matches:	125	
Percent Similarity:	50.13%	Conservative:	73	
Best Local Similarity:	31.65%	Mismatches:	139	
Query Match:	11.15%	Indels:	59	
DB:	1	Gaps:	12	
US-10-030-330-1 (1-843) x	STRSPEBH (1-1357)			
Oy	26	LysGlnArgAlaSerLeuAlaArgGluAlaLeuArgValSerLeuArgMetGly		45
Db	267	AAGAAGCAAAAGATAGCGCTATTCACATTATCCAAAATCGCA-GCATATCAAAAGCAGT		325
Oy	46	GlnThrAlaValSerAspLysIleSerLeuAspTyrValTyrArgGlnGlyAspAlaGlu		65

```

Db      326 GCACGAGCGCAGAGAT---ATTACCTTGACAAATTTACCTAGGTGGAGACTT--- 379
Qy      66 ArgGlyIleThrSerGlnGluGlySerProAlaTyrPheTyrValAlaAsnArgGly 85
Db      380 -----TCGCGCTTAATGTATGTATGTATACATATTTCTACTGGA 418
Qy      86 AsnAsnGluGlyTyrAlaLeuValAlaAlaAspAspArgIleProThrIleLeuAlaTyr 105
Db      419 -----GGATTGTGTTATCGTTTCAGAGATTAACGTTCTCCAGAAATTCATGATAC 469
Qy      106 SerProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMetTrpLeuGln 125
Db      470 TCTACACGCGGATCATTTGACGCTAACGCT---AAGAAATAATGCTTCTTCATGGA 526
Qy      126 IleTyr-----AspGlnGluIleGlyLeuIleLeuSerGlyLysAla 139
Db      527 AGTTAATGTCGACAAATCAAAAGAAACAAAATAAGTACACTTATATGCTGCTACCGCT 586
Qy      140 GlnLeuAsnGluGluIleLeuArgThrGlnGlyValProAlaGluValHisAlaLeuMet 159
Db      587 GACATTAAACACACAGTT-----GTTAATCTCTCCCT 619
Qy      160 AspAsnGlyHisPheAlaAsnAspPrometaArgTrpAsnGlnGlyTyrProTrpAsnAsn 179
Db      620 GAT-----TCAAAAGCATTCATCAATCAATCAAGTTAACCTTACAAACCTA 664
Qy      180 LysGluProLeuLeuProAsn-----GlyAsnHisAla 190
Db      665 TTGACACCTGTTATTGAAAAAGTAACAGGTAACAAATCTTTGTAGTCAACATGCA 724
Qy      191 TyrThrGlyCysValAlaThrAlaAlaGlnIleMetArgTyrHisSerTrpProLeu 210
Db      725 GCTACAGGATGTGTCTCTACTGCAACTGCTCAATATGATATCATATTAACCTTAA 784
Qy      211 GlnGluGluGlySerPheAspTyr-----HisAla 220
Db      785 AAAGGGGTGAAGACTACACTTACACACTAGCTCAATATTCATTTCAACATCCT 844
Qy      221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetTyrAspTrpIleAsn--- 239
Db      845 AAGAACTTG-----TTGCACTATCTCTACTAGCAATACAACTGGAAACAACATC 895
Qy      240 MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThr 259
Db      896 CTACCTACTTATAGCGAAGACATCTTAACGTTCAAAAAATG-----GCCATTTCAGAA 949
Qy      260 LeuMetArgAspValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThr 279
Db      950 TTGATGGCTGATGTTGATTTTCACTAGACATGATTAATGTCATCTAGTGTCTGCA 1009
Qy      280 TyrSerValTyrValAlaGlyAlaLeuArgAsnAsnPheArgTyrGlyArgSerLeuGln 299
Db      1010 GTACGCTCTCGTCTCAAGAGCCTTGAAGAAACTTTGGCTACAAACCATCTGTTCA 1069
Qy      300 LeuHisValArgAlaLeuTyrThrSerGlnGluTrpHisAspMetIleArgGlyLeu 319
Db      1070 CAATCAACCGTAGGACTTTAGCAACACAGATTGGGAGACCAAAATTTGACAAAGAAATTA 1129
Qy      320 AlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheVal 339
Db      1130 TCTCAAAACCAACGATTTACTACCAAGTGTGCGTAAGTAGGCGGACATGCTTGT 1189
Qy      340 CysAspGlyTyrAlaSerAspGlyThrPheHisPheAsnTrpGlyTrpGlyValSer 359
Db      1190 ATCGATGTGTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1249
Qy      360 AsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGly 379
Db      1250 GACGCGCTTCTTCGCTTGTAGCAGCACTTAACCTTACCTTGTGCTGCTGCGGCGCA 1309
Qy      380 IleGlyPheThrIleTyrGlnGluIleIleThrGlyIleGluPro 394

```

Db 1310 GCGGCGTTCAACGCTTACCAAAAGTGTGTGTAGGCATCAAACT 1354
 Search completed: June 2, 2003, 00:13:31
 Job time : 5689 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 1, 2003, 22:27:31 ; Search time 469 seconds

(without alignments)
4047.835 Million cell updates/sec

Title: US-10-030-330-1

Perfect score: 4438
Sequence: 1 MKKSLALVLMFGIAMQGH.....PNCATILKVDGTTTINIVH 843

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Command line parameters:

-MODE=frame_p2n.model -DEV=xlh
-O=/cgn2.1/USPTO_spool/US10030330/runat_23052003_181643.6744/app_query.fasta_1.1031
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=20
-MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030330.ecgn.1.1.290.etrnat_23052003_181643.6744 -NCPU=6 -ICPU=3
-NO_XLPTX -NO_MMAP -LANG=EUENT -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEROUT=120 -WARN_TIMEROUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_101002:*

1	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4438	100.0	2529	20	AAK91703	Porphyromonas ging
2	4438	100.0	2532	21	AAK67909	Porphyromonas ging
3	4438	100.0	2547	20	AAK91578	Porphyromonas ging
4	493	11.1	1194	24	ABN70210	Streptococcus poly
5	484	10.9	1197	17	AAK45219	Streptococcus pyog
6	484	10.9	1197	17	AAK15294	S. pyogenes spee e
7	484	10.9	1197	21	AAK07111	S. pyogenes spee e
8	331.5	7.5	415	21	AAK82217	N. meningitidis pa
9	190	4.3	1620	20	AAK91791	Porphyromonas ging
10	190	4.3	1689	20	AAK91648	Porphyromonas ging
11	164	3.7	2463	20	AAK91597	Porphyromonas ging
12	164	3.7	2607	20	AAK91572	Porphyromonas ging
13	159.5	3.6	2763	20	AAK91740	Porphyromonas ging
14	159.5	3.6	2766	20	AAK91739	Porphyromonas ging
15	159.5	3.6	2775	20	AAK91738	Porphyromonas ging
16	159.5	3.6	2814	20	AAK91610	Porphyromonas ging
17	154	3.5	3978	22	AAH81312	Escherichia coli p
18	148.5	3.3	4059	22	AAH75059	Nucleotide sequenc
19	148.5	3.3	9047	20	AAK12949	Enterococcus faeca
20	148.5	3.3	9466	23	AAK75934	DNA encoding novel
21	147.5	3.3	3413	20	AAK72330	F. dalsuclum CP70
22	144.5	3.3	4986	24	AAK37774	Alpha-tubulin
23	144	3.2	1163020	24	ABO67197	Listeria innocua c
24	144	3.2	2944528	24	ABA03041	Listeria monocylog
25	144	3.2	3011208	24	ABO69245	Listeria innocua d
26	143	3.2	1170	20	AAK91662	Porphyromonas ging
27	143	3.2	1173	20	AAK91541	Porphyromonas ging
28	143	3.2	1725	24	ABG69158	Listeria monocylog
29	143	3.2	2421	24	ABG70811	Listeria monocylog
30	142.5	3.2	1830121	17	AAK42063	Haemophilus influenzae
31	142.5	3.2	2823	24	ABO69232	Listeria sp DNA se
32	140	3.2	3106	18	AAK99905	Enterobacteriaceae
33	140	3.2	6159	22	AAK59104	M. catarrhalis str
34	140	3.2	30349	24	ABN90859	Staphylococcus epi
35	139.5	3.1	2848	18	AAK86089	Gene encoding KEX2
36	139.5	3.1	3091	18	AAK90387	Yeast KEX2 gene.
37	139.5	3.1	4106	16	AAK098696	Muramidase release
38	139.5	3.1	4118	13	AAK029472	DNA encoding novel
39	139.5	3.1	9475	23	AAK68878	Blifidobacterium bi
40	139	3.1	5509	24	ABN92640	LKPI operon. Haem
41	139	3.1	9432	17	AAK09312	Mycobacterium gall
42	138.5	3.1	2809	16	AAK51531	H. pylori HPN165 e
43	138.5	3.1	8709	22	AAK25628	Enterohaemorrhagic
44	138	3.1	3106	18	AAK02323	Protease gene. Py
45	138	3.1	4765	17	AAK08132	

ALIGNMENTS

RESULT 1	AAK91703	
ID	AAK91703	standard; DNA: 2529 BP.
XX	AAK91703;	
AC	25-AUG-1999	(first entry)
DT		
DE	Porphyromonas gingivalis protein PG28 encoding DNA.	
XX		
XX	Porphyromonas gingivalis; PG: periodontal disease; gingivitis;	
KW	vacuine; antigenic; ds.	
XX		
OS	Porphyromonas gingivalis.	
XX		
PN	W09929870-A1.	
XX		
PD	17-JUN-1999.	

XX 10-DEC-1998: 98W0-AU01023.
 XX 04-AUG-1998: 98AU-0005028.
 PR 10-DEC-1997: 97AU-0000839.
 PR 31-DEC-1997: 97AU-0001182.
 PR 30-JAN-1998: 98AU-0001546.
 PR 10-MAR-1998: 98AU-0002264.
 PR 09-APR-1998: 98AU-0002911.
 PR 23-APR-1998: 98AU-0003128.
 PR 05-MAY-1998: 98AU-0003338.
 PR 22-MAY-1998: 98AU-0003654.
 PR 29-JUL-1998: 98AU-0004917.
 XX (CSLC-) CSL LTD.
 PI Aglus CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
 PI Ross BC, Rothel LJ, Webb EA;

WPI: 1999-385613/32.
 P-PSDB: AAY34485.

PT Antigenic Porphyromonas gingivalis peptides for preventing
 PT gingivitis
 PS Claim 12: Page 209-210; 588bp; English.

CC AAY91536 to AAY91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
 CC AAY34381. AAY91802 to AAY91869 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.

SQ Sequence 2529 BP; 722 A; 561 C; 608 G; 638 T; 0 other;

Alignment Scores:
 Pred. No.: 0 Length: 2529
 Score: 4438.00 Matches: 843
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0

10-030-330-1 (1-843) x AAY91703 (1-2529)

1 MetLysLysSerPheLeuLeuAlaIleValMetLeuPheGlyIleAlaMetGlnGlyHis 20
 1 ATGAAAAAAGTTTCTTTTACCAATGTAATGCTCTTGGCATTCATGACAGGACAT 60
 21 SerAlaProValThrIleGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgIleVal 40
 61 TCTGCTCCGCTTACGAAAGAGGAGCTTGTGAGCTGGCTGGCTTGGCAGCGGTA 120
 41 SerLeuArgMetGlyGlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArg 60
 121 TCCCTTGGCAATGGGCAACACAGCAGTATGTGACAGAGATTTCATCATTCATTCG 180
 61 GlnGlyAspAlaGluArgGlyIleThrSerGlnGluGluGlySerProAlaTyrPheTyr 80
 181 CAAGGAGATGCTGAGAGGGGCTATCACATCAAGAGGAGGCTCTCCATATTTTAT 240
 81 ValAlaAsnArgGlyAsnAsnGluGlyTyrAlaLeuValAlaAlaAspArgIlePro 100
 241 GTACCTAATCTGTAATATATGAGGCTATGCTTGTGAGCAGCATATGACAGAAATCCG 300
 101 ThrIleLeuAlaTyrSerProIleGlyArgPheAspMetCaspSerMetProAspAlaLeu 120
 301 ACAATTTTAGCCTAATTCACCCATGTGGCGCTTTCGACATGACATGATGCGGCAATCTT 360

121 ArgMetTrpLeuGlnIleTyrAspGlnGluIleGlyLeuIleLeuSerGlyLysAlaGln 140
 361 CGCATGTGGCTACAAATTTACGATCGAAGAAATGGCTATACTTCCGGAAGCTCAG 420
 141 LeuAsnGluGluIleLeuArgThrGlnGlyValProAlaGluValHisAlaLeuMetAsp 160
 421 CTCATGAGAGATATATGATGATCCGAGGCGCTACCGCTGAAAGTACATCATCTCGAGGAT 480
 161 AsnGlyHisPheAlaAsnAspProMetArgTyrPheAsnGlnGlyTyrProTrpAsnLeu 180
 481 AACGGTCATTTTGGCCACATCCCATCGATGGAATCAAGATTACCATGGAACATTAAG 540
 541 GAACCATGCTTCCATATGCAATCATGCTATACGGCTGTGTCTCTCTGTCGAGCA 600
 201 GlnIleMetArgTyrHisSerTrpProLeuGlnGlyGlySerPheAspTyrHisAla 220
 601 CAATCATCGCTACCATGCTGCGCCCTTCAAGGTGAAGGCTCTTTCATATATCATGCA 660
 221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetTyrAspTrpIleAsnMet 240
 661 GGTTCATTTAGTTGGCAACTGTCCGACATTTGGGAATGTACGACTGATCATATATG 720
 241 ProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThrLeu 260
 721 CCGGAAATCCGACCTTATATCATGACTCAATCATGATGATGATGATGATGATGATGAT 780
 261 MetArgAspValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThrTyr 280
 781 ATGCGGAGATGTAGTCACTGTTTTCGAGAGATTTTATGAAGATGGAATGTGTACTAC 840
 281 SerValTyrValAlaGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGlnLeu 300
 841 AGCGTTATGTATGATGAGGACCTTTCGAAACACTTTCCTCAAGCGTTCACAGCTA 900
 301 HisValArgAlaLeuTyrThrSerGlnGluTrpHisAspMetIleArgIleGluLeuAla 320
 901 CATGTACGCGCTTATATACCTACACAGAGTGGACAGATATGATGATGATGATGATGAT 960
 321 SerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCys 340
 961 TCCGGAAGGCGGCTATATATGAGGAAATACCAAGATGAGACATCTTTCGTTTGC 1020
 341 AspGlyTyrAlaSerAspGlyThrPheHisPheAsnTrpGlyTyrGlyIleValSerAsn 360
 1021 GATGCTTATGCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
 361 GlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyIleGlyIle 380
 1081 GCTTCTACAACTAACACTACCTCTCCGACCTTCTGGGATCGAGGTGAGGGAATA 1140
 381 GlyPheThrIleTyrGlnGluIleIleThrGlyIleGluProAlaLysThrProAlaGlu 400
 1141 GGTTTTACCATTTATTAAGAGATCATCACCGGATGAGACGGCTTAAGCTCCCGTGAA 1200
 401 AlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGluAlaGlyTyrLysSer 420
 1201 GCGGATGAGATGCTTGGCGATCTTGGCAGTGAAGAAAGATGAAAGCGAGATAAAT 1260
 421 GlnSerGlyLeuAsnValIleTyrSerIleTyrAsnThrGlyGluGluGlnSerAsnLeu 440
 1261 GAATCCGAGATGAACCTAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
 441 AspLeuGlyTyrArgLeuAsnLysAlaAspGlyGluValIleGluValLysThrSerSer 460
 1321 GACCTCGATACAGATGAAACAGGCTGACGGAAGATCATAGAGTGAAGAACTCATCT 1380
 461 IleAsnIleSerTrpTyrGlyTyrGlyLysIleProGluLysPheSerLeuAlaProAsn 480
 1381 ATCATATCTCTTGTGATGAGATGAGAGCATCCGAGAGTTTCTCATTTGCACTTAAT 1440
 481 GlnLeuSerGlnGlyIleAsnThrIleThrLeuLeuTyrArgArgThrGlyThrGluGln 500

```

|||||
1441 CAGTTGTCACAGAAATCAACACACCTACTTATTCGCGCAGCAGCCGAAACG 1500
QY 501 TPGLIUPROVALARHISALAGLNGLYTYRVALASERILEYVALASNTHT 520
DB 1501 TGGGAGCCGGTAGCGCAGCAGAGAGATATGCAATGAGATTAATAATACGACA 1560
QY 521 ASPROASNAVALVALATHRVALASPAASNGLGLYLSLEUSERILEVALPRO 540
DB 1561 GACCCGAAACATGTGTACTACGCTAGTATTAACGAGAGGCAAGCTCGATCCGCC 1620
QY 541 ASNSERPHEVALALASPLEUASNSERTYRGLUHSISERTHTLETHRVALGLN 560
DB 1621 AACACCTTTGTCGAGATCTGAATCTTATGAACAATAGACATTAACATGACAT 1680
QY 561 SERASPSERPROASPGULIETHRGTHRPROVALALAPHEALAEUSERTHGLY 580
DB 1681 AGCGACACCCCTGAGAGAGATCCGTACACCCGTAGCCCTTGCTATCTACAGAG 1740
QY 581 ALAASPAPVALIILESERLEUGLYTRPVALMETLAGLVALPROGLYGLYSER 600
DB 1741 GCGGACGATGTATATCTTTGGCGTGGGTAAGGCTGAAGTTCCGGGCGGTAGC 1800
QY 601 TYRPROVALVALTRPSERLYSASPVALLEUTHRLEUSERGLUGLYASPTYR 620
DB 1801 TATCCGTTGGTTGGTCTAAAGACGTTCTCACTCTCGAAGGCGCATATACAT 1860
QY 621 TYRARGPHERILEASASNGLNLYSASPGULUTRPHYSLYLLEGLYSERVAL 640
DB 1861 TATAGATTTTCCATCAACACCAAAAGATGATGAAAGAAAATGCGAAGCGTGC 1920
QY 641 LYSTHRPROTHGLUTYRTHRHSISPROLEUPHEGLUVALGLYHSASNGL 660
DB 1921 AAAACCGCAGAGATATACGACACCCCTTATTCGAAAGTGGCCATATACAT 1980
QY 661 TYRTHRLEUASPMETALHISASNAARYALLLEUPROASPHETHRLEULYSAN 680
DB 1981 TATACGTGATATGACACACACAGAGATATGCCGACTTACACTCAAAATCG 2040
QY 681 LEUPROPHASNGLGLULEUVALVALAPHEAGLINTHGLISERSESGLYSER 700
DB 2041 TTGCTTTCAAGGAGTGTGGTGTGTGTTCCGCAACCAATCCCATCGGGG 2100
QY 701 LEUTRPAALAGLNGLUTHRVALHISILEYSGLNGLYGLUTHRPHRVALTYR 720
DB 2101 TTATGCGAGCTCAGAAACAGTACATATCAAGCAAGGAGAACTTCTATAT 2160
QY 721 VALVALGLUGLYPROILEPROASPGLYSETYRARGALATHRLEUHSIALAP 740
DB 2161 GTTGTGGAAGGCCCTATACCTGATGATCCGTATCGTGACCCCTCATG 2220
QY 741 GLYGLNGLNGLLEUTHRLEULYSGLYLSATGAENTYRTHVALYSLLEVAL 760
DB 2221 GGACAAACACAGTGTGTACTCTCAAGGGGAAAGAACTCACGTTAGATGT 2280
QY 761 THRVALVALGLUALAILLEUSERSERGLUGLILEAGVALPHRPROASN 780
DB 2281 ACAGCGGTAGAAAGCAATAGATCGTCAGAGAGATCAGATTTCCCTAAT 2340
QY 781 ASPTYRVALGLULESERLAPROCYSILEPROGLNGLUTHRSERIELLEUP 800
DB 2341 GATATGTGGAATATCGGACCTTGATCCCAAAACATCATATCTTTTCGAT 2400
QY 801 LEUSERGLYLYLLEVALMETLYSASNSERLEUSERALAGLYHISGLYARG 820
DB 2401 CTGTAGGCAAGATGTATGATGAAGATATGTTATCAGCGGGCATAGGCA 2460
QY 821 SERATGLEUPROASNGLYVALATYRILELEULYSVALASPGLYTYRTHR 840
DB 2461 AGCGACACTTCTTAATGGGGCTACATCTTAAGGTGATGATATACGAGAA 2520
QY 841 ILEVALHIS 843
|||||

```

```

DB 2521 ATAGTCAC 2529
RESULT 2
ID AAC67909 standard; DNA: 2532 BP.
AC AAC67909;
DT 16-FEB-2001 (first entry)
XX
XX Porphyromonas gingivalis periodontal.
DE Porphyromonas gingivalis; periodontal; antiinflammatory; antibacterial;
XX Porphyromonas gingivalis; periodontal; antiinflammatory; antibacterial;
KM amidolytic; alpha-1-proteinase inhibitor; periodontitis; gingivitis; ds.
XX
XX Porphyromonas gingivalis.
OS
XX
XX W0200063394-A2.
XX
XX 26-OCT-2000.
XX
XX 20-APR-2000; 2000MO-US10574.
XX
XX 21-APR-1999; 99US-0130436.
XX
XX (UYGE-) UNIV GEORGIA RES FOUND INC.
PA (TRAY/) TRAVIS J.
PA (POTE/) POTEMPA J.
PA (NELS/) NELSON D.
XX
XX Travis J, Potempa J, Nelson D;
PI
XX
XX WPI: 2000-679600/66.
DR P-PSDB; AAB36102.
XX
XX Novel oral bacterial periodontal polypeptide for treating periodontal
PT diseases, has amidolytic activity for cleavage of non-denatured human
PT alpha1-proteinase inhibitor at reactive site loop region of inhibitor
PT
XX
XX Claim 24; Fig 5; 55pp; English.
XX
XX The present sequence is given in a specification relating to novel
CC oral bacterial polypeptide referred to as periodontal. The polypeptide
CC has amidolytic activity for cleavage of denatured polypeptides and
CC non-denatured serpin polypeptides. It has amidolytic activity for
CC cleavage of a non-denatured human alpha-1-proteinase inhibitor at a
CC reactive site loop region of the inhibitor. Periodontal is useful for
CC inhibiting the peptidase activity and reducing periodontitis, loss of
CC tooth attachment and periodontal pocket formation, and for reducing
CC growth of bacteria, preferably P. gingivalis in vitro or in vivo.
CC It is useful for protecting an animal from a disease caused by
CC P. gingivalis and for treating periodontal diseases, including
CC gingivitis and periodontitis.
XX
XX Sequence 2532 BP; 724 A; 561 C; 608 G; 639 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 0 Length: 2532
Score: 4438.00 Matches: 843
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0
US-10-030-330-1 (1-843) x AAC67909 (1-2532)
QY 1 MetLysLysSerPheLeuLeuAlaIleValMetLeuPheGlyTyrLeaMetGlnGlyHis 20
DB 1 ATGAAAAAAGTTTCTTTTACCATGATATGCTTTTGGCATTCCTCCATGCGAGACAT 60
QY 21 SerAlaProValThrLysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgGlnVal 40
|||||

```

```

Db      61  TCTGCTCGGTTACGAAGAAGCGAGCTTGTGAGTCTGGCTCGGCTGCTTCCGACAGCTA 120
Qy      41  SerLeuArgMetCylGlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArg 60
Db      121  TCCCTCCGAATGGGAAACACAGCATCTGCACAAAGATTCCATTCGATTACGTTATCGG 180
Qy      61  GlnGlyAspAlaGlnArgGlyIlePheSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 180
Db      181  CAAGGAGATGCTGAGAGGGGTATCATCATCAACAGAGAGAGCTCTCCGATTTTAT 240
Qy      81  ValAlaAsnArgGlyAsnAsnGlnGlyTyrAlaLeuValAlaAlaAspAspArgIlePro 100
Db      241  GTAGCTATCTGGGAAATATAGGGCTATGCTCTGTGACACACATGACAGAAATACCG 300
Qy      101  ThrIleLeuAlaTyrSerProIleGlyArgPheAspMetAspSerMetProAspAsnLeu 120
Db      301  ACAATTTTACGCTTATCCACCATTTGGCCGTTTGACATGACAGATATGCCGACAACTT 360
Qy      121  ArgMetTyrLeuGlnIleTyrAspGlnGlnIleGlyLeuIleLeuSerGlyValAlaGln 140
Db      361  CGCATGTGGCTACAAATTTACGATCAGAAATAGCCGTGATCTTCCGAAAAGCTCAG 420
Qy      141  LeuAsnGlnGlnIleLeuArgThrGlnGlyValProAlaGlnValHisAlaLeuMetAsp 160
Db      421  CTCAATGAAAGATATTACGTACCGAGGCGTACCGGCTGAGAGTACATGCTGTGATGAT 480
Qy      161  AsnGlyHisPheAlaAsnAspProMetArgTyrPheAsnGlnGlyTyrProTyrPheAsnLys 180
Db      481  AACGGTCAATTTGGCCACGATCCCATGCGATGAGATCAAGTATCCCATGGAACAAATTAAG 540
Qy      181  GluProLeuLeuProAsnGlyAsnHisAlaTyrThrGlyValAlaThrAlaAlaAla 200
Db      541  GAAACACGCTCTCCATATGCAATCATGCTATACCGCGTGTCTCTACTGCTGACGCA 600
Qy      201  GlnIleMetArgTyrHisSerTyrProLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 600
Db      601  CAATTCATGCGCTACCATACGTGCGCGCTTCAAGAGTGAAGGCTCTTTCATATATCATCA 660
Qy      221  GlySerLeuValGlyAsnTyrSerGlyIlePheGlyValMetTyrAspTyrIleAsnMet 240
Db      661  GCTTCATTAATGTTGGCAACTGCTGCGCACATTTGGTGAATGTATACGATCATATATG 720
Qy      241  ProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAlaAlaTyrAlaThrLeu 260
Db      721  CCGGGAATTCGCCACCTGTATATCTCAATCTCAATGCGATGCTACGCGCACACTG 780
Qy      261  MetArgAspValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThrTyr 280
Db      781  ATGCGTGAATGATGATGATGCTGTTTCATGATGATTTTATGAAATGGAATGGATGGTAC 840
Qy      281  SerValTyrValValGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGlnLeu 300
Db      841  AGCGTTTATGTATGATGAGACCTTGCGAACAACTTTCGTTACAAAGCTTCACTGCACTGA 900
Qy      301  HisValAlaGlnAlaLeuTyrThrSerGlnGlnIlePheAspMetIleArgIleLeuAla 320
Db      901  CATGTACGCGCTTATATCTCATCCACAGAGTGCACAGATATGATCCGCGGGAACCTTGC 960
Qy      321  SerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCys 340
Db      961  TCCGGAAGGCGGCTATTTATGCAAGGAAATACACAGACATGAGACATGCTTCTGTTGC 1020
Qy      341  AspGlyTyrAlaSerAspLysThrPheHisPheAsnTyrPglTyrPglGlyValSerAsn 360
Db      1021  GATGCTTATGCTTCCGATGCTGATCTTCCATTTCAACTGGGTTGGGAGGTTTCCAC 1080
Qy      361  GlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlnGlyLys 380
Db      1081  GGCCTTCTCAAACTAACACTCTCTCCGCGCACTTCTGTGGTATCCGAGGTGAGGAAATA 1140
Qy      381  GlyPheThrIleTyrGlnGlnIleIleThrGlyIleGlnProAlaIleThrProAlaGln 400
Db      1141  GATTTCACATTTATCAAGAGATCATCACCGGTATGGAACCGGCTTAAGACTCCCGCTGA 1200

```

```

Qy      401  AlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGlnAlaGlnIleTyrLysSer 420
Db      1201  GCCGTTACAGATGCTTCCGATCTTGGCACTGCAAGACATGAAAGCCAGATATAAAGT 1260
Qy      421  GluSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGlyGlnGlnGlnGlnGlnGln 1260
Db      1261  GAATCCGGATGAAACGATAGGCTATTCGATATATATACGGGGAAGCAATCAAACTCT 1320
Qy      441  AspLeuGlyTyrArgLeuAsnLysAlaAspGlyValIleGlyValLysThrSerSer 460
Db      1321  GACCTCGATACAGATGTAACAAAGCTGACGAGAGATCAATAGAGGGAAGTCAAACTCA 1380
Qy      461  IleAsnIleSerTyrTyrGlyIleGlyLysIleAspProGlySerPheSerLeuAlaProAsn 480
Db      1381  ATCAATATCTCTGTGTACGATACGAGAGCATCCGAGACTTCTCATTTGGCACTTAAT 1440
Qy      481  GlnLeuSerGlnGlyIleAsnThrIleThrLeuLeuTyrArgTyrThrGlnGln 1500
Db      1441  CAGTTGTCAAGAAATCAACACCATCACTTATCTGTCGACAGCACCCAGCAAG 1500
Qy      501  TrpGluProValArgHisAlaGlnGlyIleTyrValAsnSerIleLysValAsnThr 520
Db      1501  TGGAGCGCGGTACCGCATGCAAGGGGATATGTCATATAGCATTTAAAGTAAATACGCA 1560
Qy      521  AspProAsnAsnValValIleThrValAspAsnGlnGlyLysLeuSerIleValPro 540
Db      1561  GACCCGAAATGCTGATGATCAAGATTAATTAAGAAAGCAACCTCATATATGCTGCC 1620
Qy      541  AsnSerPheValAlaAspLeuAsnSerTyrGlnHisSerThrIleThrValGlnPheAsn 560
Db      1621  AACAGCTTGTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
Qy      561  SerAspSerProAspGlnIleArgThrProValAlaPheAlaLeuSerThrGlyAlaThr 580
Db      1681  AGGACAGCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
Qy      581  AlaAspAspValIleSerLeuGlyTyrValMetAlaGlnValProGlyLysSerAsn 600
Db      1741  GCGACGATGATGATATCTTGGCTGATGATGATGATGATGATGATGATGATGATGAT 1800
Qy      601  TyrProValValTyrSerLysAspValLeuThrLeuSerGlnGlyAspTyrThrLeuPro 620
Db      1801  TATCCGCGGTTGCTGTTAAAGCTTCTCACTCTCCGAAAGCGGACATATCATTTGG 1860
Qy      621  TyrArgPheSerIleAsnAsnGlnLysAspGlnTyrLysIleGlySerValSerVal 640
Db      1861  TATGATTTTCTCAATCAACCAACCAAAAGATGATGAAAGAAAGATGGAAGGTGTCA 1920
Qy      641  LysThrProThrGlnIleTyrThrHisProLeuPheGlnValGlyHisAsnGlnThrSerThr 660
Db      1921  AAAACACCGCAGAGATACGACCCCTTATTCGAAGGGGCCCAATCAACTTTCAC 1980
Qy      661  TyrThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLysAsnLeuGly 680
Db      1981  TATACCTGATATGATGACACACAGATATTTGCCGACCTTTACACTCAAAATATCCGA 2040
Qy      681  LeuProPheAsnGlyLysLeuValValPheArgGlnThrGlnSerSerGlySer 700
Db      2041  TTGCTCTTCAATGAGGATGTTGTTGTTTCCGCAACCAACCAACTTCATCGGGGTCT 2100
Qy      701  LeuThrAlaAlaGlnGlnIleThrValHisIleLysGlnGlyIleThrPheValTyrLysPro 720
Db      2101  TTATGGGAGCTCAAGAACAGATACATCAACCAAGAGAAACTTTCGATATTAACCT 2160
Qy      721  ValValGlnGlyProIleProAspGlySerTyrArgAlaThrLeuHisAlaPheValAsn 740
Db      2161  GTTGTGAAAGCCCTATACCTGATGATCTTATCTGCAACCTTCACAGCATTTGTGTAAC 2220
Qy      741  GlyGlnGlnGlnLeuTyrLeuLysGlyLysArgAsnTyrThrValLysIleValAsnGly 760
Db      2221  GGACAAACAAGTGTACTCAAGGGGAAAGAACTACAGGTGAAGATCTCATATGCT 2280

```

QY 761 ThrAlaValGluAlaIleGluSerSerGluGluIleArgValPheProAsnProAlaArg 780
DB 2281 ACAGGGGTAGAGCAATGAAATCGTCAGAGATCAGAGTATTCCTTAATCCGGCAGCC 2340
QY 781 AspTyrValGluIleSerAlaProCysIleProGluIleThrSerIleIleLeuPheAsp 800
DB 2341 GATTATGCGAAATACCGCACCCTTGCAATCCCAAGAACATCATTCATTCTTTTCGAT 2400
QY 801 LeuSerGlyLysIleValMetLysAsnSerLeuSerAlaGlyHisGlyArgMetAspVal 820
DB 2401 CTGTCAGGCAAGATTCATGATGAGATAGTTATCAGCGGGGCATGCGCAATGGAATGTC 2460
QY 821 SerArgLeuProAsnGlyAlaTyrIleLeuLysValAspGlyTyrThrThrLysIleAsn 840
DB 2461 AGCGGACTTCCTTAATGGGGCCTACATCTTAAGTGATGATATACGAGAAATTAAT 2520
QY 841 IleValHis 843
2521 ATAGTGCAC 2529
AT 3
1578
ID AAX91578 standard; DNA; 2547 BP.
XX AAX91578;
AC AAX91578;
AT 25-AUG-1999 (first entry)
DE Porphyromonas gingivalis protein PG28 ORF encoding DNA.
XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
KW vaccine; antigenic; ds.
XX Porphyromonas gingivalis.
OS Porphyromonas gingivalis.
PN W09929870-A1.
PD 17-JUN-1999.
XX 10-DEC-1998; 98MO-AU01023.
PF 10-DEC-1998;
XX 04-AUG-1998; 98AU-0005028.
PR 10-DEC-1997; 97AU-0000839.
PR 31-DEC-1997; 97AU-0001182.
PR 30-JAN-1998; 98AU-0001545.
PR 10-MAR-1998; 98AU-0002264.
PR 09-APR-1998; 98AU-0002911.
PR 23-APR-1998; 98AU-0003128.
05-MAY-1998; 98AU-0003338.
22-MAY-1998; 98AU-0003654.
29-JUL-1998; 98AU-0004917.
XX (CSLIC-) CSL LTD.
PA Aglus CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX MPI: 1999-385613/32.
DR P-RSDB; AAY34360.
XX Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis
PS Claim 12; Page 124; 588pp; English.
XX AAY91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC assays. Porphyromonas gingivalis is involved in periodontal disease

CC especially gingivitis.
XX Sequence 2547 BP; 727 A; 566 C; 610 G; 644 T; 0 other;
SQ Alignment Scores:
Pred. No.: 0 Length: 2547
Score: 4438.00 Matches: 843
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 20 Gaps: 0
US-10-030-330-1 (1-843) x AAX91578 (1-2547)
QY 1 MetLysLysSerPheLeuLeuAlaIleValMetLeuPheGlyIleAlaMetGlnGlyHis 20
DB 19 ATGAAAAAAGTTTCTTTTACCATGTAATGCTTTTGGCAATTCCTTCATGCGGACAT 78
QY 21 SerAlaProValThrLysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgGlnVal 40
DB 79 TCTGCTCCGGTTACGAAAGAGGAGCTTGAGCTGCTGCGCTTGCGACAGTA 138
QY 41 SerLeuArgMetGlyGlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArg 60
DB 139 TCCTTGCGAATGGGCAAAACACAGCATCTGCACAAAGATTCCATTCATTCATTCG 198
QY 61 GlnGlyAspAlaGluArgGlyIleThrSerGlnGluGlySerProAlaTyrPheTyr 80
DB 199 CAAGGAGATCGAGGGGGGTATCACATCACAAGAGAGGCTCTCCATATTTTAT 258
QY 81 ValAlaAsnArgGlyAsnAsnGluGlyTyrAlaLeuValAlaAlaAspArgIlePro 100
DB 259 GTAGCTATCTCGGAATATATAGGGCTTGTCTGTACGACACATGACAGAAATACCG 318
QY 101 ThrIleLeuAlaTyrSerProIleGlyArgPheAspMetAspSerMetProAsnLeu 120
DB 319 ACAATTTTAGCCTATTCACCCATTGGCCCTTCGACATGACAGTATGCGGCAATCTT 378
QY 121 ArgMetTrpLeuGlnIleTyrAspGlnGluIleGlyLeuIleLeuSerGlyLysAlaGln 140
DB 379 CGCATGTGGCTCAAAATTTAGCATGCAAGAAATAGGCTCATATCTTCGGAATGCTCAG 438
QY 141 LeuAsnGluGluIleLeuArgThrGlnGlyValProAlaGluValHisAlaLeuMetAsp 160
DB 439 CTCATGAAAGATATATACGTACCGAGGCGTCACGCTGACACTACATGCTCGATGAT 498
QY 161 AsnGlyHisPheAlaAsnAspProMetArgTyrPasnGlnGlyTyrProTyrPasnAsnLys 180
DB 499 AACGGTCATTTTGGCAACGATCCCATGGAATCAAGTTACCATGGAAACAATAAG 558
QY 181 GluProLeuLeuProAsnGlnLysAsnHisAlaTyrThrGlyCysValAlaThrAlaAla 200
DB 559 GAACCACTGCTCTTAATGCAATCAATGCTTACCGCTGTCTCTACTCTGTCAGCA 618
QY 201 GlnIleMetArgTyrHisSerTrpProLeuGlnGlyGlySerPheAspTyrHisAla 220
DB 619 CAATATCATCGCTACATACGTGCGCTTCAAGGTGAAGGCTCTTCATTTATCATGCA 678
QY 221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetTyrAspTrpIleAsnMet 240
DB 679 GGTTCATTTAGTTGGCAACGTGTCGCGCATATTTGGTAATGACAGCATGATATAG 738
QY 241 ProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAlaAspAlaTyrAlaThrLeu 260
DB 739 CCCGGAATCCGACCTTGATATCTGATCAATCAAGTGAATCCATACGCGCACATG 798
QY 261 MetArgAspValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThrTyr 280
DB 799 ATGCGTGAATGATGATGATCTGTTGATGATGATTTTATGAATGATGATGATGATGAT 858
QY 281 SerValTyrValValGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGlnLeu 300
DB 859 ACGTTTATGTAGTAGGACCTTGGCAACACTTTCGTACAAAGGCTTCACTGACGCTA 918

OY 301 HisValArgAlaLeuTyrThrSerGlnGluTrpHisAspMetCIIArgGlyGluLeuAla 320
 DB |||||||
 OY 919 CATGTACGGCCCTTAAATACCTCACAGAGTGGCCACGATGATATCCGGGGAAACTTGCC 978
 DB |||||||
 OY 321 SerGlyArgProValTyrTyrAlaGlySAsnAsnGlnSerIleGlyHisAlaPheValCys 340
 DB |||||||
 OY 979 TCGGAAAGCCCGCTATTATCGCAGGAAATTAACAGACATAGACATAGCATGCTTCGTTTC 1038
 DB |||||||
 OY 341 AspGlyTyrAlaSerAspGlyThrPheHisPheAsnTrpGlyTyrPglYglValSerAsn 360
 DB |||||||
 OY 1039 GATGGTTATGCTTGGATGTGACTTCCATTTCACATGGGGTGGGGAGGTGTTTCCAAC 1098
 DB |||||||
 OY 361 GlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGluGlyIle 380
 DB |||||||
 OY 1099 GGCTTCTACAAACATCAACACTCTCTCCGCCATCTCTGGGTATCGGAGCTGAGGGAATA 1158
 DB |||||||
 OY 381 GlyPheThrIleTyrGlnGluIleIleThrGlyIleGluProAlaLysThrProAlaGlu 400
 DB |||||||
 OY 1159 GCTTTTACATTTATCAAGAGATCATCCCGGTATCGAACCGGCTTACACTCCCGTGA 1218
 DB |||||||
 OY 401 AlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGluAlaGlyTyrLysSer 420
 DB |||||||
 OY 1219 GCCGCTACAGATCCCTTGGCGATCTTGGCAGTGAAGACATAGAGCCGAGTATTAAGT 1278
 DB |||||||
 OY 421 GluSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGlyGluGlnSerAsnLeu 440
 DB |||||||
 OY 1279 GAATCCGATGTAACGTAGGATTCGATATATATACAGGTGAAAGAGCAATCAATCTT 1338
 DB |||||||
 OY 441 AspLeuGlyTyrArgLeuAsnLysAlaAspGlyIleValIleGluValLysThrSerSer 460
 DB |||||||
 OY 1339 GACCTCGATACAGATTATACAAAGCGTACGAGGAGAACTCATAGAGTGAACACTTATCT 1398
 DB |||||||
 OY 461 IleAsnIleSerTrpTyrGlyTyrGlyLysIleProGluSerPheSerLeuAlaProAsn 480
 DB |||||||
 OY 1399 ATCATATCTCTTGTGTCAGATACGAGAGACATCCCGAAGTTTCTCATTTGGCACCTAT 1458
 DB |||||||
 OY 481 GlnLeuSerGlnGlyIleAsnThrIleThrLeuTyrArgArgThrGlyThrGlnGln 500
 DB |||||||
 OY 1459 CAGTTGTCACAGGAATCAACACCATCCCTTACTTATCTGCGACAGGACCGAACAG 1518
 DB |||||||
 OY 501 TrpGluProValArgHisAlaGlnGlyTyrValAsnSerIleLysValAsnThrThr 520
 DB |||||||
 OY 1519 TCGGAGCCGCTACGCGATCGACAGAGAGATATGTCATATACATTAAAGTAAATACACA 1578
 DB |||||||
 OY 521 AspProAsnAsnValValIleThrValAspAsnAsnGlnGlyLysLeuSerIleValPro 540
 DB |||||||
 OY 1579 GACCCGAACAATGTCGTAGTCACGCTAGATATATACAGAGCAAGCTATATCGTCCC 1638
 DB |||||||
 OY 541 AsnSerPheValAlaAspLeuAsnSerTyrGluHisSerThrIleThrValGlnPheAsn 560
 DB |||||||
 OY 1639 AACACCTTGTGCGAGATCTGAATCTTATGAACATGTAGCATTTACATACAGTTCAAT 1698
 DB |||||||
 OY 561 SerAspSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrGlyAlaThr 580
 DB |||||||
 OY 1699 AGCCAGACCCCTGTATGATCGCTACACCGGTACCTTGTCTCTACAGAGCACTACT 1758
 DB |||||||
 OY 581 AlaAspAspValIleSerLeuGlyTyrPheLysAlaGluValProGlyGlySerSerAsn 600
 DB |||||||
 OY 1759 GCGAGCATGTATATCTTGGCTGGTAAATGCTGAAGTTCCGGGCGGTAGCCAGCAAC 1818
 DB |||||||
 OY 601 TyrProValValTrpSerLysAspValLeuThrLeuSerGluGlyAspTyrThrLeuTrp 620
 DB |||||||
 OY 1819 TATCCGCTGTTGGTCTAAAGACTTCTCACTCTCTCGGAAGCGACTTATCATTTGTGG 1878
 DB |||||||
 OY 621 TyrArgPheSerIleAsnAsnGlnLysAspGluTrpLysLysIleGlySerValSerVal 640
 DB |||||||
 OY 1879 TATAGATTTCATCAACAAACAAAGATGAAGAAAAAGATCGAAGCGGTCTAGTA 1938
 DB |||||||
 OY 641 LysThrProThrGlyTyrThrHisProLeuPheGluValGlyHisAsnGlnThrSerThr 660
 DB |||||||
 OY 1939 AAAACACCGACAGATACGACACCCCTTATCTGGAAGTGGCCATATCAAACTTCTAC 1998
 DB |||||||

OY 661 TyrThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLysAsnLeuGly 680
 DB |||||||
 OY 1999 TATACCGGTATATGGACACAAACAGATGTTTCCCGACTTTCACCTCAAAAATCTCCGA 2058
 DB |||||||
 OY 681 LeuProPheAsnGlyGluLeuValValPheArgGluThrGlnLeuSerSerGlySer 700
 DB |||||||
 OY 2059 TTGCTTTCAATGGTGTGATGTTGTTGTTTCCGCCAAACAACTCTCATCGGGGCT 2118
 DB |||||||
 OY 701 LeuTrpAlaAlaGlnGluThrValHisIleLysGlnGlyGluThrPheValTyrLysPro 720
 DB |||||||
 OY 2119 TTTATGGGACCTTCAAAACAGTATCATATCAAGCAAGAGAACTTCTGATATTAACCT 2178
 DB |||||||
 OY 721 ValValGluGlyProIleProAspGlySerTyrArgAlaThrLeuHisAlaPheValAsn 740
 DB |||||||
 OY 2179 GTTGTGAAAGGCCCTATACCTGATGATCCATGCTCGACCCCTCATTCATTCGTAAAC 2238
 DB |||||||
 OY 741 GlyGlnGlnGluLeuTyrLeuLysGlyLysArgAsnTyrThrValLysIleValAsnGly 760
 DB |||||||
 OY 2239 GGACAAACAAGTGTGTACTTCAAGGGGAAAGAACTACACGGTGAAGATCTCATATGT 2298
 DB |||||||
 OY 761 ThrAlaValAlaGluAlaIleGluSerSerGluGluIleArgValPheProAsnProAlaArg 780
 DB |||||||
 OY 2299 ACAGCGGTAGAAAGAAATGATGTCAGAAAGATCAGATATCCCTTAATCCGGCACGC 2358
 DB |||||||
 OY 781 AspTyrValGluIleSerAlaProCysIleProGlnGluThrSerIleIleLeuPheAsp 800
 DB |||||||
 OY 2359 GATTATGTGAATATTCGGCACCTTCATTCCTCCCAAGAAACATCATTCATCTTTGAT 2418
 DB |||||||
 OY 801 LeuSerGlyLysIleValMetLysAsnSerLeuSerAlaGlyHisGlyArgMetAspVal 820
 DB |||||||
 OY 2419 CTGTAGGCAAGATTTCTATGAAAGATGTTATACGGGGCATGGCAAAATGAGATCT 2478
 DB |||||||
 OY 821 SerArgLeuProAsnGlyAlaTyrIleLeuLysValAspGlyTyrThrLysIleAsn 840
 DB |||||||
 OY 2479 AGCCGACTTCTTAAGGGGCGCTTACATCTTAAAGTGTGATGATACGAGAAATTAAT 2538
 DB |||||||
 OY 841 IleValHis 843
 DB |||||||
 OY 2539 ATAGTGAC 2547
 DB |||||||

RESULT 4
 ABN70210
 ID ABN70210 standard; DNA; 1194 BP.
 AC ABN70210;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE Streptococcus polynucleotide SPO ID NO 8333.
 XX
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 OS antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
 XX
 PN Streptococcus pyogenes.
 PN
 PD WO200234771-A2.
 PD
 XX 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-GB04789.
 PF
 XX 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 PR
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 XX Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C;
 PI Tettelin H;
 XX
 DR MPI; 2002-352536/38.

PF 30-APR-1996; 96WO-US05997.
 XX
 PR 01-MAY-1995; 95US-0432692.
 XX
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX
 PI Ananthaswamy HN, Fernandez A, Kapur V, Musser JM;
 XX
 DR WPI: 1996-506148/50.
 DR P-PSDB; AAM07898.
 XX
 PT Use of extracellular Streptococcal cysteine protease enzyme - for
 PT inhibiting the proliferation of neoplastic cells, e.g. for treating
 PT carcinoma, lymphoma or leukemia.
 XX
 PS Disclosure: Pages 62-63; 99pp; English.

The present sequence encodes the streptococcus pyogenes clone speB7
 pre-pro cysteine protease (CP), which can be used to inhibit
 neoplastic cell proliferation, especially in a human, useful in the
 treatment of neoplastic conditions, e.g. carcinomas, sarcomas,
 melanomas, lymphomas and leukemias originating from blood, lung,
 mammary gland, prostate, intestine, stomach, liver, heart, skin,
 pancreas or brain tissue. The CP is especially associated with a
 wound covering, and can also be used to prevent metastasis or
 CC identify susceptible neoplastic cells.
 CC K1735 and CM519 melanoma cells were injected s.c. into nu/nu mice,
 CC optionally followed by i.p. injection of CP (100 microg, 24 hours
 CC later). The mice were checked twice weekly for tumor growth for
 CC 12 weeks, to give results that showed that treatment with CP
 CC completely protected athymic mice against transplanted K1735
 CC melanoma growth, and protected 60% of the mice from developing
 CC CM519 melanoma.

SO Sequence 1197 BP; 393 A; 239 C; 235 G; 330 T; 0 other:

Alignment Scores:

Pred. No.:	1-79e-31	Length:	1197
Score:	484.00	Matches:	125
Percent Similarity:	49.62%	Conservative:	72
Best Local Similarity:	31.49%	Mismatches:	138
Query Match:	10.91%	Gaps:	63
	17		13

US-10-030-330-1 (1-843) x AAT45219 (1-1197)

```

26  LysGluAlaGlnLeuSerLeuAlaArgLeuAlaLeuAlaGlnValSerLeuArgMetGly 45
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107  AAGAGGCAAAAGATGCGCTATCATTCATTCACAAAATACAGCA-GCTATCAAAACAGGT 165
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46  GlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArgGlnGluAspAlaGlu 65
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
166  GCACGACGCCGAGAAAT-ATTAAAGCTTGACAAAGTTAACTTAGGTGGAGAACTT- 219
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
66  ArgGlyIleThrSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 85
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
220  -----TCTGCGCTTAATATGTTATGTTACAAATATTCTACTGGA 258
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
86  AsnaGlnGlnGlyTyrAlaLeuValAlaAlaAspAspArgIleProThrIleLeuAlaTyr 105
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
259  -----GGAATTTGTTATCGTTTCAGAGATAACGTTCTCCAGAAATCTTAGATAC 309
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
106  SerProIleGlyArgPheAspMetAspSerMetProAspAsnIleArgMetIlePleuGln 125
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
310  TCTACACGCGATCTTTGACGCTTAACGT-AAAAGAAACATTCCTTCTTACGAA 366
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
126  IleTyr-----AspGlnGlnIleGlyLeuIleLeuSerGlyTyrAla 139
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
367  AGTATGTGCAACAATCAAGAAACAAATAATTAGACACTTATGCTGCTGACCGGT 426
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
140  GlnLeuAsnGlnGlnIleLeuArgThrGlnGlyValProAlaGlnValHisAlaLeuMet 159
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  
```

```

Db 427  GAGATTAAACACAGTT-----GTTAAATCTCTCTT 459
Oy 160  AspaGlnGlnHisPheAlaAsnAspProMetLArgTrpAsnGlnGlyTyrProTrpAsn 179
Db 460  GAT-----TCAAAAGCATTCATTACACCAAGGATTAACCTTACACCTA 504
Oy 180  LysGluProLeuLeuProAsn-----GlyAsnHisAla 190
Db 505  TTGACACCTGTTATGAAAGGTAAGAACAGGTGAAACATCTTTGTAGTCAACATCA 564
Oy 191  TyrThrGlyCysValAlaIleThrAlaAlaIleGlnIleMetLArgTyrHisSerTrpLeu 210
Db 565  GCTACAGGATGTGTGCTACTGCAACCTCTCAATATGAATATCAATATTAACCT- 621
Oy 211  GlnGlyGlnGlySerPheAspTyr----- 218
Db 622  ---AACAAAGGTTGAAAGACTTACACTTACACACTTACCAATTAACCATTTTCAAC 678
Oy 219  HisAlaGlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetTyrAspTrpIle 238
Db 679  CATCTAGAACTTG-----TTGACAGTATCTCTACTAGACAAATACAGTGAAC 729
Oy 239  Asn---MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyr 257
Db 730  AACATCTTACTTACTTATAGCGGAGAGAAATCTTAAGTCAAAAATG- -GCGAT 783
Oy 258  AlaThrLeuMetArgAspValSerValSerValSerMetSerPheTyrGluAsnGlySer 277
Db 784  TCAGATTTGATGCGCGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 843
Oy 278  GlyThrTyrSerValTyrValAlaGlnAlaLeuArgAsnAsnPheArgTyrLysArgSer 297
Db 844  TCTGACGATGCTCTGCTGTTCAAGACCTTGAAGAAACTTTGGCTACACCAACT 903
Oy 298  LeuGlnLeuHisValArgAlaLeuTyrThrSerGlnGlnGlnGlnGlnGlnGlnGlnGln 317
Db 904  GTTCACCAAAATTAACCGTACGCTTATGCAAAAGATGGAGAACCAAAATTAACAA 963
Oy 318  GluLeuAlaSerGlyArgProValTyrTyrAlaGlnGlnGlnGlnGlnGlnGlnGlnGln 337
Db 964  GAATATTCGAAACCAACCAACGATATACGACGATGCGGTAAGTGGCGGACATGCC 1023
Oy 338  PheValCysAspGlyTyrAlaSerAspGlyThrPheHisPheAsnTrpGlyTyrPleuGly 357
Db 1024  TTTGTTATGACGATGCTGCTGACGAGCTTAACCTTACATGTTAACTGGGTTGGGTGGA 1083
Oy 358  ValSerAsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGly 377
Db 1084  GTCTGTGACGGCTTCTTCGTTGACGCACTAAACCTTCAGCTTGGTACTGTGGTGC 1143
Oy 378  GlnGlyIleGlyPheThrIleTyrGlnGlnIleIleThrGlyIleGluPro 394
Db 1144  GGGCGAGCGGCTTCAACGCTTACCAAGTCTGTGTAGGACATTAACCT 1194
  
```

RESULT 6
 AAT15294
 ID AAT15294 standard; cDNA; 1197 BP.
 XX
 AC AAT15294;
 XX
 DT 30-OCT-1996 (first entry)
 XX
 DE S. pyogenes speB extracellular protease gene.
 XX
 KW Immunogenic peptide: speB gene; extracellular protease; production;
 KW antibody; vaccine; diagnosis; detection; Streptococcus infection;
 KW group A; prevention; treatment; pharyngitis; tonsillitis;
 KW skin infection; acute rheumatic fever; scarlet fever; probe;
 KW post-streptococcal glomerulonephritis; sepsis; meningitis;
 KW erysipelas; cellulitis; fasciitis; toxic shock like syndrome; ss.
 XX
 OS Streptococcus pyogenes.
 XX

Key Location/Qualifiers
 1.1197
 /tag=a
 /transl_except= pos:646..648, aa:Tyr
 W09608569-A2.
 21-MAR-1996.
 13-SEP-1995; 95MO-US11723.
 14-SEP-1994; 94US-0306542.
 (BAYLOR COLLEGE MEDICINE.
 Kapur V, Musser JA;
 WPI. 1996-179944/18.
 P-PSDB; AAR95856.

Use of extracellular protease(s), partic. cysteine protease - for
 detection, diagnosis prevention and treatment of infection by
 pathogenic organisms, partic. gp. A streptococcus strains
 Disclosure: Page 18; 97pp; English.

The present sequence is the S. pyogenes speB extracellular protease
 gene. An immunogenic peptide derived from the protease can be
 used in the prodn. of antibody (Ab) and vaccine. Ab is prepd. by
 introducing the peptide into a mammal, pref. a mouse, followed
 by Ab isolation. The Ab or a probe derived from the gene can be
 used for the diagnosis and detection of gp. A Streptococcus
 infections, while vaccine, which inhibits streptococcal
 infections, can be used to prevent and treat gp. A Streptococcus
 infections, and partic. to ameliorate pharyngitis, tonsillitis,
 skin infections, acute rheumatic fever, scarlet fever,
 post-streptococcal glomerulonephritis, sepsis, meningitis,
 erysipelas, cellulitis, fasciitis and toxic shock like syndrome.

Sequence 1197 BP: 393 A; 239 G; 235 G; 330 T; 0 other;

Alignment Scores:
 Pred. No.: 1.79e-31 Length: 1197
 Score: 484.00 Matches: 125
 Percent Similarity: 49.62% Conservative: 72
 Best Local Similarity: 31.49% Mismatches: 138
 Query Match: 10.91% Indels: 63
 Gaps: 13

D-030-330-1 (1-843) x AAT15294 (1-1197)

26 LysGIuAlaGlaLeuSerLeuAlaArgLeuAlaLeuArgGlnValSerLeuArgMetGly 45
 107 AAGAAGCAAAAGATGCGCTATCATTCACAAATAACACA -GCTATCAAAAGCAGGT 165
 46 GlnThrAlaValSerAspLysIleSerIleAspTyrValTyrArgGlnGlyAspAlaGlu 65
 166 GCACGACGCGCAGAGAT---ATTAGCTTGACAAAGATTAACTTACGTGAGAACTG--- 219
 66 ArgGlyIleThrSerGlnGluGlySerProAlaTyrPheTyrValAlaAsnArgGly 85
 220 -----TCTGCGCTTAATATGATGATGTTACAAATTTCTACTGGA 258
 86 AsnAsnGluGlyTyrAlaLeuValAlaAlaAspArgGlyIleProThrIleLeuAlaTyr 105
 239 -----GCAATTTGTTATCGTTTCAGAGATTAACGCTCCGAAATTTCTAGCATAC 309
 106 SerProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMetTyrLeuGln 125
 310 TCTACACAGCGATCATTTGACCTTAACGCT---AAAGAAACATATGCTTCTCATGCA 366
 126 IleTyr-----AspGlnGluIleGlyLeuIleLeuSerGlyLysAla 139

367 AGTTATGTGACAAATCAAGAAACAAATAATAGACACTATTAGTGTACCGCT 426
 140 GlnLeuAsnGluGluIleLeuArgThrGluGlyValProAlaGluValHisAlaLeuMet 159
 427 GAGATTAACCAACCACTT-----GTTAAATCTCTCCTT 459
 160 AspAsnGlyHisPheHisAsnAspPrometArgTTPAsnGlnGlyTyrProThrAsnAsn 179
 460 GAT-----TCAAAAGCATTCATTACACCAAGTAAACCTTACAACTTA 504
 180 LysGluProLeuLeuProAsn-----GlyAsnHisAla 190
 505 TTGACACACTGTATTATAAAAGTAAACCAAGTGAACCAATCTTTGTAGTCAACATGCA 564
 191 TyrThrGlyCysValAlaThrAlaAlaGlnIleMetArgTyrHisSerTyrProLeu 210
 565 GCTACAGATGTGTCTCTACGTGCAACTGCTCAAAATTAATGATAATATTCATTCCT--- 621
 211 GlnGlyGluIleSerPheAspTyr----- 218
 622 ---AACAAAGGTTGAAGACTTACACTTACACTTAACTCAATACCAATATTCATTCAC 678
 219 HisAlaGlySerLeuValGlyAsnTyrSerGlyThrPheGlyGluMetTyrAspTyrIle 238
 679 CATCCTAAGACACTTG-----TTGACAGTATCTCTACATGACAAATACATGCAAC 729
 239 Asn---MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyr 257
 730 AACATCTCTCACTTATTAAGCGGACAGATCAATCAACAAATG-----CGAATT 783
 258 AlaThrLeuMetArgPheValSerAlaSerValSerMetSerPheTyrGluAsnGlySer 277
 784 TCAAGATTATAGTGTATGTTGATTTTCAAGATGATGATGATGATGATGATGATGATGAT 843
 278 GlyThrTyrSerValTyrValGlyAlaLeuAlaGlyAsnAsnPheArgTyrLysArgSer 297
 844 TGTGACAGTATGCTGTGCTGCTCAAAAGACCTTGAAGAAACCTTGGCTACACCAATCT 903
 298 LeuGlnLeuHisValAlaGluAlaLeuTyrThrSerGlnGluThrPheAsnMetIleArgGly 317
 904 GTTCACCAAAATTAACCGTATGACGCTTATGACAAACAAAGATGGAGACCAAAATTAACA 963
 318 GlnLeuAlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlnHisAla 337
 964 GAATTATCTCAAAACCAACAGATTAATCAACAAAGTGTGGTAAAGTGGCGACATGCC 1023
 338 PheValCysAspGlyTyrAlaSerAspGlyThrPheHisPheAsnThrPglTyrGlyGly 357
 1024 TTTGTTATGATGATGCTGACGACGCTTACTTCTACATGTTAACTGCGGTGGGTGGA 1083
 358 ValSerAsnGlyPheTyrTyrLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGly 377
 1084 GTCTCTGACGCGCTCTTCGCTTGACGACATTAACCCCTTCTGCTGCTGCTGCTGCTGCT 1143
 378 GlnGlyIleGlyPheThrIleTyrGlnGluIleThrGlyIleGlyPro 394
 1144 GCGCAGCGCGCTTCAACGCTTACCAAAAGTGTGTTGAGGCACTAAACT 1194

RESULT 7
 AAA07111 standard; cDNA; 1197 BP.
 AAA07111;
 09-JUN-2000 (first entry)

S. pyogenes cysteine protease speB coding sequence.
 Cysteine protease; speB. Group A streptococcus; extracellular protease;
 detection; diagnosis; extracellular matrix; infection; skin infection;
 disease status monitoring; vaccine; Streptococcus mediated disease;
 pharyngitis; tonsillitis; scarlet fever; sepsis; erysipelas; fasciitis;
 pneumonia; acute rheumatic fever; poststreptococcal glomerulonephritis;

QY 136 SerGlyLysAlaGlnLeuAsnGluGluIleLeuArg-----ThrGluGlyValPro 152

CC used as vaccines especially

CC	used as vaccines especially against Porphyromonas gingivalis. Probes can
CC	activity with a vaccine mechanism of action. The PG polypeptides have antibacterial
CC	isolation of the PG polypeptides. The PG polypeptides represent PCR primers used in the
CC	AAY3483. AAX91802 to AAX91989 represent PCR primers used in the
CC	Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC	AAX91536 to AAX91801 encode two hundred and sixty six antigenic
XX	Claim 12; Page 270; 588pp; English.
PS	
PX	
PT	
PI	Antigenic Porphyromonas gingivalis peptides for preventing
PI	gingivitis
PA	(CSLC-) CSL LTD.
PR	29-JUL-1998; 98AU-0004917.
PR	22-MAY-1998; 98AU-0003654.
PR	05-MAY-1998; 98AU-0003338.
PR	23-APR-1998; 98AU-0003128.
PR	09-APR-1998; 98AU-0002911.
PR	10-MAR-1998; 98AU-0002264.
PR	30-JAN-1998; 98AU-0001546.
PR	31-DEC-1997; 97AU-0001182.
PR	10-DEC-1997; 97AU-0000839.
PR	04-AUG-1998; 98AU-0005028.
PF	10-DEC-1998; 98WO-AU01023.
PD	
PN	17-JUN-1999.
PO	
OS	W09929870-A1.
RK	Porphyromonas gingivalis.
RN	
RQ	
RW	Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
RZ	vaccine; antigenic; ds.
SX	
TA	
TB	
TC	
TD	
TE	
TF	
TG	
TH	
TI	
TJ	
TK	
TL	
TM	
TN	
TO	
TP	
TR	
TS	
TT	
TU	
TV	
TW	
TX	
TY	
UZ	
VV	
WW	
XX	
YY	
ZZ	
AA	
AB	
AC	
AD	
AE	
AF	
AG	
AH	
AI	
AJ	
AK	
AL	
AM	
AN	
AO	
AP	
AQ	
AR	
AS	
AT	
AV	
AW	
AX	
AY	
AZ	
BA	
BB	
BC	
BD	
BE	
BF	
BG	
BH	
BI	
BJ	
BK	
BL	
BM	
BN	
BO	
BP	
BQ	
BR	
BS	
BT	
BU	
BV	
BW	
BX	
BY	
BZ	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CZ	
DA	
DB	
DC	
DD	
DE	
DF	
DG	
DH	
DI	
DJ	
DK	
DL	
DM	
DN	
DO	
DP	
DQ	
DR	
DS	
DT	
DU	
DV	
DW	
DX	
DY	
DZ	
EA	
EB	
EC	
ED	
EE	
EF	
EG	
EH	
EI	
EJ	
EK	
EL	
EM	
EN	
EO	
EP	
EQ	
ER	
ES	
ET	
EU	
EV	
EW	
EX	
EY	
EZ	
FA	
FB	
FC	
FD	
FE	
FF	
FG	
FH	
FI	
FJ	
FK	
FL	
FM	
FN	
FO	
FP	
FQ	
FR	
FS	
FT	
FU	
FV	
FW	
FX	
FY	
FZ	
GA	
GB	
GC	
GD	
GE	
GF	
GG	
GH	
GI	
GJ	
GK	
GL	
GM	
GN	
GO	
GP	
GQ	
GR	
GS	
GT	
GU	

CC be used to detect Porphyromonas gingivalis in standard hybridisation
assays. Porphyromonas gingivalis is involved in periodontal disease
especially gingivitis.

XX Sequence 1620 BP; 478 A; 346 C; 410 G; 386 T; 0 other;

Alignment Scores:

Pred. No.:	1,93e-06	Length:	1620
Score:	190.00	Matches:	103
Percent Similarity:	35.98%	Conservative:	78
Best Local Similarity:	20.48%	Mismatches:	162
Query Match:	4.28%	Indels:	160
DB:	20	Gaps:	23

US-10-030-330-1 (1-843) x AAX91791 (1-1620)

```

OY 413 APTLIEGLALGLUTYRYSERGLYLEUASNAVAL----- 426
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 301 GATATCTGCTGTAACGGGTAAAGATGAAATCCGACATCAAGATTGGTGGTAGAGCTCATG 360
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 427 -----GlyTyr-----SerileTyrAsnThrGlyGluGluGlnSer 438
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 361 AATAAGCCCGGAGATATAGAGTAGAGTGGGCTTCAGTCGGGATCCCAACGGCGAG 420
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 439 ASNLLEUASPLEUGLYTYRARGLEUASNYLSALAASPGLYGLUVALILEGLUVALYSTR 458
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 421 AATGGCAAACTGCTGATATAGGAAAGACTTCCCAATGTCAGATGACATGCTGATATA 480
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 459 SERSEILEASNILESERTPYRGLYRGLYGLUHLSPROGLUSERPHESEILEU 478
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 481 GCCTCC-----AACTAT-----CGTTCG 498
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 479 PROASGLNLEUSERGLNYLLEASNTHRIETHRLEUETRYARGRTHRGLYTHR 498
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 499 CCTTCTCTCTTAACAATGGTGGCAACCTTTGGCTTGGCTTGGCTTACACGGCTTC 558
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 499 GUGINTTPGLUPROVALARGHISALAGLNGLYTYRVALASNSERILEUYSVALSN 518
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 558 ----- 558
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 519 THRTHRAPPROASNAVALVALVATHRVALASNAASGLUYLSEUSERILE 538
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 559 -----AACATACACCAAAATTAAGTTT 582
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 539 VALPROASNSERPHEVALALAASPLEUASNSERTYRGLUHLISERTHRILETHRYALGN 558
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 583 GTG-----GACTATGTGTCTCTGATGAGGCAAAATTTCAATAAAGCTTACAC 636
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 559 PHEASNSERASPSERPROASPLULEARGTHRPROVALALAPHEALAEUSERTHGLY 578
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 637 TTCAGTCAAGATGAGAGAAATTTGAC-----ACGTGATCTCTCATTTGGT 687
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 579 ALATHRALASAPVALILESERLEUGLYTRPVALMETALGLUVAL----- 594
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 688 AGCAGCTCGATCCATGCGTCAACAATGCTGCGCCGCTAATGGGTGTGATTCCAAATG 747
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 595 -----PROGLYGLYSERASANTYRPROVAL----- 604
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 748 AATAAACAAGGGGGAAGGATATGCGTTCTTGTGCAACTTGTGCAATGATGCC 807
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 605 -----TTPSER-----LYS 607
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 808 GAATTCAGTGTGTCAGCCCTATAAAGTACGTGAAGCGACATGCTGTCAGCCCAAA 867
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 608 ASPVALLEUTHLEUSERGLUYASPTYRTHLEU-----TRPYRARGPHE 623
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 868 ATCCAAATGTTCTGTGACAGCATATACCAATACGATCAATGGGAGAGTTGCCACAATTC 927
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 624 SERILE-----ASNAASGLNLSASPGLYTRPLYSILIEGLYSERVAL 638
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 928 ATGATTACGTACAGCATATGATCTGTAATATGCGATGCG----- 969
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 639 SERVALYSTRTPROTHRGILUTYRTHRHSIPROLEUPHEGLUVALGLYHISASGLNTHR 658

```

```

DB 970 -----GACATTCGATGATATATCCACAGAAATCGTCAAGATGAAAAAGGA 1017
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 659 SERTHRTRYRTHLEUASPMETALHISASNAVALLEUPROASPHE-----THRLEU 676
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1018 AAACCTCCGACTATGATGATCTGTGGAAGCTTCTTACAGCTTCGACCAAGTGAG 1077
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 677 LYSASNLLEUGLYLEUPROHE-----ASNGLYGLUVALVALVATHRARG 692
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1078 ACCAACTCGGGCTGGGATACAGAAAGCCCAATCACTACCTATACATATGCCAAA 1137
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 693 GINTHRGINSERSEGLYSERLEU-----TRPALALAGLNUTHRVALHIS 709
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1138 AAGAAGAGAACGTCAGCAACGCTGAATACCGCTGGCCCAATTAACAAAGATCAT 1197
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 710 ILEYSGLN-----GLYLURPHE-----VALTYRYS 719
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1138 AACAAAGATTGTGGAGCGACATTTACGTTATCATCATCTGCCAATGCTCTTACACA 1257
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 720 PROVALVALGLUGLYPROLLEPROASPLYSERTYRARGALATHRLEUHLAPHEVAL 739
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1258 CCTCAAGTAGAC-----ATC 1272
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 740 ASNGLYGLNGLNLEU-----TYRLEUYSGLYLSARGASN 752
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1273 AATCCGACCAAGGCTGTGTGTGTCATGATGATGATATATCCCGGCAAAAGGATC 1332
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 753 TYR-----THVALYSILEVALASNGLYTHRALVALGLUVALILEGLUSERGLU 770
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1333 GTTGGCTGATACGACGAGGACCATGCCAAGCGGTGTAAGACATGTAATGCAGAA 1392
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 771 ---GLUILEARYVALPHEPROASNPROALARGASPTYRVALGLUILESERALAPROCY 789
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1393 GGCACATCAAGCTCTACCCGAATCCGCTCAAGAATATGCTGATATAGCTGCCG--- 1449
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 790 ILEPROGLNUTHRSERILELEUPHEASPLEUSERGLYLSILEVALMETYLSASN 809
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1450 ACGGACCAAACTGCAAGGCTGTGTTACGATATGCAAGGAGAGATGATGCTGAGCT 1509
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 810 SERLEUSERALAGLYHISGLYARGMETASPVALLSERARGLEUPROASGLYVALTYRILE 829
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1510 TCTTCTCTCGGCAAGATATACGCTGAAACGTGCAAGCTTGGTAAAGGTACATACATA 1569
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 830 LEUYSVAL 832
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1570 CTCAGGTC 1578
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 10
AAX91648 standard; DNA; 1689 BP.
AAX91648;
25-AUG-1999 (first entry)
Porphyromonas gingivalis protein PG91 ORF encoding DNA.
Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
vaccine; antigenic; ds.
Porphyromonas gingivalis.
W09929870-A1.
17-JUN-1999.
10-DEC-1998; 98MO-AU01023.
04-AUG-1998; 98AU-0005028.
10-DEC-1997; 97AU-0000839.
31-DEC-1997; 97AU-0001182.
30-JAN-1998; 98AU-0001546.
10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.
 PR 23-APR-1998; 98AU-0003128.
 PR 05-MAY-1998; 98AU-0003338.
 PR 22-MAY-1998; 98AU-0003654.
 PR 29-JUL-1998; 98AU-0004917.

XX (CSLC-) CSL LTD.

PI Agus CT, Bart IG, Hocking DM, Margetts MB, Patterson MA;
 PI Ross BC, Rothel LJ, Webb EA;

DR WPI: 1999-385613/32.
 DR P-PSDB: AAY34430.

XX Antigenic Porphyromonas gingivalis peptides for preventing
 PT gingivitis
 XX

PS Claim 12, Page 172; 588bp; English.

AAK91536 to AAK91801 encode two hundred and sixty six antigenic
 Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
 AAY34553. AAK91802 to AAK91989 represent PCR primers used in the
 isolation of the PG polypeptides. The PG polypeptides have antibacterial
 activity with a vaccine mechanism of action. The PG polypeptides can be
 used as vaccines especially against Porphyromonas gingivalis. Probes can
 be used to detect Porphyromonas gingivalis in standard hybridisation
 assays. Porphyromonas gingivalis is involved in periodontal disease
 especially gingivitis.

XX Sequence 1689 BP; 505 A; 361 C; 423 G; 400 T; 0 other;

Alignment Scores:

Pred. No.: 2,04e-06 Length: 1689
 Score: 190.00 Matches: 103
 Percent Similarity: 35.98% Conservative: 78
 Best Local Similarity: 20.48% Mismatches: 162
 Query Match: 4.28% Indels: 160
 DB: 20 Gaps: 23

US-10-030-330-1 (1-843) x AAK91648 (1-1689)

QY 413 AspIleGluAlaGluTyrLysSerGluSerGlyLeuAsnVal----- 426
 DB 370 GATATCGTGTAAACGGTAAAGATGATCCGACATCAAGATTGGTCGTAGAGCTCATG 429
 QY 427 -----GlyTyr-----SerIleTyrAsnThrGlyGluGluGlnSer 438
 DB 430 AATAAGCCCGGAGGATATAGAGTAGAGTTGGGCTTCAGTCGCGATGCCAAGCGCGAG 489
 QY 439 AsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyGluValIleGluValLysThr 458
 DB 490 AATGGGAACATCGTGTATAGGAAGACATTCCTCAATGTCGACATTCAGATTCGATATA 549
 QY 459 SerSerIleAsnIleSerTyrGlyGlyGluHisProGluSerPheSerLeuAla 478
 DB 550 GCTCTC-----AACTAT-----CGTTCG 567
 QY 479 ProAsnGlnLeuSerGlnGlyLysAsnThrIleThrLeuLeuTyrArgArgThrGlyThr 498
 DB 568 CCTTCTCTCTTAACAAATGTCGCAACCTTTGCTTGGCTTGGCTTACACGCGCTTC 627
 QY 499 GluGlnTyrGluProValArgHisAlaGlnGlyTyrValAsnSerIleLysValAsn 518
 DB 627 ----- 627
 QY 519 ThrThrAspProAsnAsnValValThrValAspAsnAsnGluGlyLysLeuSerIle 538
 DB 628 -----AACATACGACCAAAATTAAGTTT 651
 QY 539 ValProAsnSerPheValAlaAspLeuAsnSerTyrGluHisSerThrIleThrValGln 558
 DB 652 GTG-----GACTATGTGTCTCTCTGATGAGGCGCAAAATTCATTAATAACTTACTC 705

QY 559 PheAsnSerAspSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrGly 578
 DB 706 TTCAGTCAGATGAGAGAGAAATTTGAC-----AAGTGGATCTCTCATGGGT 756
 QY 579 AlaThrAlaAspValIleSerLeuGlyTyrValMetAlaGluVal----- 594
 DB 757 AGCAGCTCTGAATCCATGGGTGCACATGCTGCGCCCTAATGGGTGGATTCGAATG 816
 QY 595 -----ProGlyGlySerSerAsnTyrProValAl----- 604
 DB 817 AATAAACAAGGGGAAAAAGCGATATCGTTCTGTGCAATTTGTCGACAAATGATCCC 876
 QY 605 -----TyrSer-----Lys 607
 DB 877 GAATTTCAGTGGTCAGAGCCCTATAAAGTACGAAAGCGACATGTCGTTACAGCCCAA 936
 QY 608 AspValLeuThrLeuSerGluGlyAspTyrThrLeu-----TyrTyrArgPhe 623
 DB 937 ATCCAAATGTGTGTCGACAGATTAACAATACGATCAATGGGAGATTGCCAACCTTC 996
 QY 624 SerIle-----AsnAsnGlnLysAspGluTyrLysIleGlySerVal 638
 DB 997 ATGATTACGTCACGCGATTATGATCTGATATTCGATTCGATTG----- 1038
 QY 639 SerValLysThrProThrGluTyrThrHisProLeuPheGluValGlyHisAsnGlnThr 658
 DB 1039 -----GACTTGGGTGTGTATATCCCAAGAAATCTTCAAGTATGAAAAAGA 1086
 QY 659 SerThrTyrThrLeuAspMetAlaHisAsnArgValLeuProAspPhe-----ThrLeu 676
 DB 1087 AAACCTCCGACATGATGATCTGTGGAAGCTTCTTCCATCGTTCACAGTACAGTACAG 1146
 QY 677 LysAsnLeuGlyLeuProPhe-----AsnGlyGluLeuValValAlaPheArg 692
 DB 1147 ACCCAATCGGGGCTGGGTGATGACAAAGAACCCCAATCACTGATTCATATGCGCAA 1206
 QY 693 GlnThrGlnSerSerSerGlySerLeu-----TyrAlaAlaGlnGluThrValHis 709
 DB 1207 AAAGAAGACACGCTGCAACACGCTGAATATCCGCTGGCCAAATATGCAAGATTCAT 1266
 QY 710 IleLysGln-----GlyGluThrPhe-----ValTyrLys 719
 DB 1267 AACCAAGATTGTGGAGCGACACATCTAGCTATACATCTGCCAATGCTCTCTACACA 1326
 QY 720 ProValValGluGlyProIleProAspGlySerTyrArgAlaThrLeuHisAlaPheVal 739
 DB 1327 CCTCAAGTAGAC-----ATC 1341
 QY 740 AsnGlyGlnGlnLeu-----TyrLeuLysGlyLysArgAsn 752
 DB 1342 AATCCGACCAAGGCTCTGTCGTCGTCATGCTGCGATATCTCCCGGCAACGCGATC 1401
 QY 753 Tyr-----ThrValLysIleValAsnGlyTyrAlaValAlaIleGluSerSerGlu 770
 DB 1402 GTTTGGTGTATACGACGATGACCATGCCAAGCGGTGAGAACATCGTAATGCAAGAA 1461
 QY 771 -----GluIleArgValPheProAsnProAlaArgAspTyrValGluIleSerAlaProCys 789
 DB 1462 GCGACGATACGCTACCCGAATTCGCTCAAGAAATATGCTGTGATACCTCGCG-- 1518
 QY 790 IleProGlnGluThrSerIleIleLeuPheAspLeuSerGlyLysIleValMetLysAsn 809
 DB 1519 ACGGCAACAACTGCAAGGCTGTGTATTACGATATCAGGCGCAGATAGCTGTAGGCT 1578
 QY 810 SerLeuSerAlaGlyHisGlyArgMetAspValSerArgLeuProAsnGlyAlaTyrIle 829
 DB 1579 TCTTCTCCGCAACGAATACAGGCTGACGCTGACACTTGGCTAAGGGTACGTACATA 1638
 QY 830 LeuLysVal 832
 DB 1639 CTCGAAGTC 1647

RESULT 11

AA91697
ID AAX91697 standard: DNA: 2463 BP.
AC AAX91697;
XX
XX
XX 25-AUG-1999 (first entry)
DE Porphyromonas gingivalis protein PG21 encoding DNA.
XX
XX Porphyromonas gingivalis; PG: periodontal disease; gingivitis;
XX vaccine; antigenic; ds.
XX
XX Porphyromonas gingivalis.
XX
XX W09929870-A1.
XX
XX 17-JUN-1999.
XX
XX 10-DEC-1998; 98MO-A001023.
XX
XX 04-AUG-1998; 98AU-0005028.
XX 10-DEC-1997; 97AU-0000839.
XX 31-DEC-1997; 97AU-000182.
XX 30-JAN-1998; 98AU-0001546.
XX 10-MAR-1998; 98AU-0002264.
XX 09-APR-1998; 98AU-0002911.
XX 23-APR-1998; 98AU-0003128.
XX 05-MAY-1998; 98AU-0003338.
XX 22-MAY-1998; 98AU-0003654.
XX 29-JUL-1998; 98AU-0004917.
XX
XX (CSLC-) CSL LTD.
XX
XX Agius CT, Barr IG, Hocking DM, Margetta MB, Patterson MA;
XX Ross BC, Rochel LJ, Webb EA;
XX WPI: 1999-385613/32.
XX P-PSDB: AAY34479.
XX
XX
XX Antigenic Porphyromonas gingivalis peptides for preventing
XX gingivitis
XX
XX Claim 12: Page 205-206; 588pp; English.
XX
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
XX Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
XX AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
XX isolation of the PG polypeptides. The PG polypeptides have antibacterial
XX activity with a vaccine mechanism of action. The PG polypeptides can be
XX used as vaccines especially against Porphyromonas gingivalis. Probes can
XX be used to detect Porphyromonas gingivalis in standard hybridisation
XX assays. Porphyromonas gingivalis is involved in periodontal disease
XX especially gingivitis.
XX
XX
XX Sequence 2463 BP: 729 A; 634 C; 565 G; 535 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 0.000535
Score: 164.00 Length: 2463
Percent Similarity: 32.498 Matches: 194
Best Local Similarity: 20.466 Conservative: 114
Query Match: 3.708 Mismatches: 293
Indels: 351
Gaps: 45
US-10-030-330-1 (1-843) x AAX91697 (1-2463)
QY 23 ProValThrLysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgGlnVal----- 40
DB 280 COTCATCATCATGACGCGCGCTGTCAGATGTT-CGACTACCGCTCAGGAGAAATACGA 338
QY 41 -----SerLeuArgMetGlyGlnThrAlaValaSerAspLys 52
DB 339 ATGAAATGCTTACCCACCATATGAGCGCTATATACACATGATGAGAGAGACTTCACAAACA 398

QY 53 IleSerIleAspTyrValTyr-----ArgGlnGly----- 62
DB 399 GTATCATCATCTTTGACTACTCTTCGTCATTTGGCAAGTCCGTAAGATCGTAACATGAT 458
QY 63 AspAlaGluArgGlyIleThrSerGln-----GluGluGlySerProAla---TyrPhe 79
DB 459 GATTTCGACGCTGACGCTCTGTCGCAATACAGGAGAAAGCTCGCTGCTTACTCTC 518
QY 80 TyrValAlaAsnArgGlyAsnAsnGluGlyTyrAlaLeuVal----- 93
DB 519 TACGATGCACGAGAGAGAAAGACAC- -GGATATGTGTGCTGCTCGCATCATAGACAT 576
QY 94 -----AlaAlaAspAspArgIleProThrIleLeuAlaTyrSerProIle 108
DB 577 CTGCTGTGCAACTACGAAATCCGATCCGAGATTAAGATTTCTG----- 621
QY 109 GlyArgPheAspMetAspSerMetProAspAsnLeuArgMetIlePheGlnIleTyrAsp 128
DB 622 -----GATTAACGGAAGTATGATGTG-CCCTTTGAC 653
QY 129 GlnGluIleGlyLeuIleLeuSerGlyLysAlaGlnLeuAsnGluGlnIleLeuArgThr 148
DB 654 CAA-----TCCGAGCGAGCATACAGACCGGAAACACACACCGTACA 695
QY 149 GluGlyValProAlaGlu-ValHisAlaLeuMetAspAsnGlnIleHisPheAlaAsnAsp- 167
DB 696 AGGAGCTACTCGCTACAAATGCCAATGTCATTTGAACCGTAACTTCAAGATGATGT 725
QY 168 -----PrometArgTrpAsnGlnGlyTyrProTrp-AsnAsnLysGluProL 183
DB 756 AGCGGTGATCACCC-----GATGAAACCTTGGCAGCCGAGGACAA- 799
QY 183 euleuProAsnGlyAsnHisAlaTyrThrGlyCysValAlaIleAlaAlaGlnIleM 203
DB 799 ----- 799
QY 203 eArgTyrHisSerTrpProLeuGlnGlyGlySerPhe-----AspTyrH 219
DB 800 -----CTGCATTCATGAT-TTGGAGAGAAACCTCTTCTGCTCGTCCCAATATAC 853
QY 219 IsAlaGlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetTyrAspTrpIleA 239
DB 854 ATGGAGGAGACAGAGTGTGAC- -TATCATG- 885
QY 239 smetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAsp- 255
DB 886 -----GATATTAATAAAGAAAGACATGCAGACGATGATGT 270
QY 256 -----AlaTyrAlaIleLeuMetAlaArgAspValSerAlaSerValSerM 270
DB 923 ACAAACTGATCAGTCGCAACTACGACCCCTTGTCCAGATATTTCCGCCAGCTACATCA 982
QY 270 eSerPheTyrGluAsnGlySerGlyTyrTrpSer-----ValTyrValValGlyA 287
DB 983 CCTCC-----GAACCAATTCGGGAATCATCAACGGTTACAGCTGTAATTAATTCGG 1036
QY 287 IalLeuArg-----AsnAsnPheArgTyrLysArgSerLeuGlnHisValA 303
DB 1037 GAATGCGTACAGCAATGCAAAATTAATTTCCATGCTGTGCGAATAATTCCTGAAATC- 1095
QY 303 rGAlaLeuTyrThrSerGlnGluTrpHisAspMetIleArgGlyGluLeuAlaSerGlyA 323
DB 1096 -----AGCAACACGAAGTGTGCGCGCTCTCAAC 1126
QY 323 rgrProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValaCysAspGlyT 343
DB 1127 TTCCAAGATGATGATGACAAAGATCTCTG----- 1161
QY 343 yrlaSerAspGlyThrPheHisPheAsnTrpGlyTyrGlyValaSerAsnGlyPhe 363
DB 1161 ----- 1161


```

Db 1708 -----GCCATGAGTACAGAGCAGAAATCCGCTGTAATCCATAGTCTATCCCGGTCATAG 1762
QY 512 aAlaSerIleuVal---AsnThrThiAspProAsnValVal-----V 527
Db 1763 ACGTTACGCTCAAGTGTGAGTGAAGTGTCCACAGATACAGAAAGAAATATCA 1822
QY 527 aIThrValAspAsnGlnGlyLysLeuSerIleValPro---AsnSerPheValAla 546
Db 1823 TCACGTGCAT-----GCCGTATGCTCTAGCTGTAATGCTGGA 1864
QY 546 sPLeuAsnSerIleGlnHisSerThrIleThrValGlnPheAsnSerAspSerProAspG 566
Db 1865 CCCCAGCGAATAGAGAGGCGCAG---ACGGTATCTTCCAAAACCAATCCACCAATG 1921
QY 566 IuIleThrThrProValAlaPheAlaLeuSerThrGlyAlaThrAlaAspAspValIle 586
Db 1922 CCACCACTACCTATGATATGATGCGGCGACCTCCCTACAGTGAAGAC----- 1974
QY 586 eLeuGlyTrpValMetAlaGluValProGlyGlySerSerAsnTrpProValValTrpS 606
Db 1975 -----GAAACCCGACTGTCCTTACA 1996
QY 606 eTrLysAspValLeuThrLeuSerGlnGlyAspTrpThrLeuTrpTrpArgPheSerIle 626
Db 1997 GCAAA-----GCCGCAATAGATGTCACGCTC----- 2025
QY 626 sNAsnGlnLysAspGluTrpLysIleGlySer-----ValSerValLysTrpProt 644
Db 2026 -----AAGCGCATAGTGTCCGCGTGAAGAGCGAAG 2065
QY 644 hGluTrpThrHisProLeuPheGluValGlyHisAsnGlnHisSerThrTrpThrLeu 664
Db 2066 AAAATATC----- 2073
QY 664 sPMeAlaHisAsnArgValLeuProAspPheThrLeuLysAsnLeuGlyLeuProPhe 684
Db 2074 -----ATCAGTGTCAAGAAAGTCGCGGCCCT----- 2100
QY 684 sNGLyGluLeuValValPheArgGlnThrGlnSerSerGlySerLeuTrpAla 704
Db 2101 -----GCTCCGCTACCGCAGGAGAACACCTCGAATA----- 2136
QY 704 IaGlnGluThrValHisIleLysGlnGlyLysThrPheValLysTrpPro----- 720
Db 2137 -----GTAAGAAAGGCGAGACGTTACTTCAAAAGCTGTGTACA 2179
QY 721 -----ValValGluGlyProIlePro----- 727
Db 2180 ACAATCCGACTCATGCGTTGGGTGTGGAAGGCGGCTCTCTGCCACACAGCGAGC 2239
QY 728 -----AspGlySerTrpArgAlaThrLeuHisAlaPheV 739
Db 2240 AAAACCCGCTGTCACCTACAAATGAACAGCAAGCATGATCCAGCTAGTCCACCA 2299
QY 739 aLAsnGlyGlnGlnGlnLeuTrpLysGlyLysArgAsnTrpThrValLysIleVal 759
Db 2300 ACGAGGGGGAAGCAAT-----GTGAAGAAAGCAGAAACATAC--ATTGAGGTATCC 2350
QY 759 sNGLyThrAlaValAlaIleGlnLeuSerSerGlnGluLeuArgValPheProAsnPro 779
Db 2351 TCGATGACAGGTGTCGACATAGGCGACAGAGCGGTATCGTCATTCGCGCAAAAG 2410
QY 779 IaArgAspTrpValGluIleSerAlaProCysIleProGlnGluThrSerIleIleLeu 799
Db 2411 GAACCAAGCAGATCTCTATGAGCAACGCTGTATCAAA-----GGCATGCTTCTCT 2464
QY 799 hAspLeuSerGlyLysIleValMetLysAsnSerLeuAlaGlyHisGlyArgMet 819
Db 2465 ATGACATCATGAGGAGGCTGACTCAAAACTCCGATCAAGCTCCGCTGACCGGTAG 2524
QY 819 sPValSerArgLeuProAsnGlnLysAlaTrpIleLeuLysVal-----AspGlyTr 835

```

```

Db 2525 ATCTTCCATCTCGCCGCAAGAAATCTACCAATCATATCAAAAGGAAATCCGCTC 2584
QY 835 yThrThrLysIleAsnIle 841
Db 2585 GCACGGAAAAAGATCATATC 2604

RESULT 13
ID AAX91740 standard; DNA: 2763 BP.
XX
AC AAX91740;
XX
DT 25-AUG-1999 (first entry)
XX
DE Porphyromonas gingivalis protein PG57 encoding DNA.
XX
KW Porphyromonas gingivalis; PG: periodontal disease; gingivitis;
XX vaccine; antigenic; ds.
XX
OS Porphyromonas gingivalis.
XX
XX MO929870-A1.
XX
PD 17-JUN-1999.
XX
PF 10-DEC-1998; 98AU-AU01023.
XX
PR 04-AUG-1998; 98AU-0005028.
XX 10-DEC-1997; 97AU-0000839.
XX 31-DEC-1997; 97AU-0001182.
XX 30-JAN-1998; 98AU-0001546.
XX 10-MAR-1998; 98AU-0002264.
XX 09-APR-1998; 98AU-0002911.
XX 23-APR-1998; 98AU-0003128.
XX 05-MAY-1998; 98AU-0003338.
XX 22-MAY-1998; 98AU-0003654.
XX 29-JUL-1998; 98AU-0004917.
XX
XX (CSLC-) CSL LTD.
XX Agius CT, Barr IG, Hocking DM, Margolis MB, Patterson MA;
XX Ross BC, Rothel LJ, Webb EA;
XX
XX WPI: 1999-385613/32.
XX P-PSDB: AAY34522.
XX
XX Antigenic Porphyromonas gingivalis peptides for preventing
XX gingivitis
XX
XX Claim 12: Page 235-236; 588pp; English.
XX
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
XX Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
XX AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
XX isolation of the PG polypeptides. The PG polypeptides have antibacterial
XX activity with a vaccine mechanism of action. The PG polypeptides can be
XX used as vaccines especially against Porphyromonas gingivalis. Probes can
XX be used to detect Porphyromonas gingivalis in standard hybridisation
XX assays. Porphyromonas gingivalis is involved in periodontal disease
XX especially gingivitis.
XX
XX SO Sequence 2763 BP; 767 A; 580 C; 666 G; 750 T; 0 other;

Alignment Scores:
Pred. No.: 0.0015 Length: 2763
Score: 159.50 Matches: 152
Percent Similarity: 31.14 Conservative: 108
Best Local Similarity: 18.208 Mismatches: 231
Query Match: 3.598 Indels: 345
DB: 20 Gaps: 38

US-10-030-330-1 (1-843) x AAX91740 (1-2763)

```

```
OY 221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyIleuMetTyraSPTrIleAsnMet 240
Db 643 GGTGAGAAATGTGGACGATGACTGTGCAACTATCCGAGGATATAT-----CAA 650
OY 241 Pro-----GlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyR 257
Db 691 CCGGAGAGAAAGGGGAATGAGAGAGTTG---CACGTTAGCGGCTACACATCTATGCGAAC 747
OY 258 AlaThrLeuMetAlaGlyAspValSerAlaSerValSerMetSerPheThrGlyIleuAsnGlySer 277
Db 748 GGTACACTACTAGGACAAATATAA-----GATGTCCTCCATCTAGGAGATATGAGAC 798
OY 278 GlyThrTySerValTyValValGlyAlaLeuArgAsnAspPheArgTyRysArgSer 297
Db 799 AGCCTACTCTCT-----TTCCGAGACAAAT-----CCC 825
OY 298 LeuGlnLeuHis-----ValArgAlaLeuTyRThrSerGlnGluTrpHisAspMetIle 315
Db 826 TTCCAGAGTGAGTACGCGCTTACAGCGCGTTTAC-----858
OY 316 ArgGlyGlnLeuAlaSerGlyArgProValTyTyValGlyAlaAsnGlnSerIleGly 335
Db 859 -----873
OY 336 HisAlaPheValCysAsp-----GlyTyValaSerAspGlyThrPheHisPheAsnTrp 353
Db 874 TCTTGACCGCTATGTGGCAGCGCTGATTCAGCCAGGATGCCATCTTATGAAATTTT 933
OY 354 GlyTPGlyGlyValaSerAsnGlyPheThrTyRysLeuThrLeuLeuSerProThrSerLeu 373
Db 934 GAGATGGACCTGTTCCCAATGCTTGC-----CTT 963
OY 374 GlyIleGlyGlyGlyIleGlyIleGlyPhe-----382
Db 964 GTGATAGACGCTGATGGAGATGATTAACCTGGGACACTATTGATGATACAGCGCT 1023
OY 383 -----ThIleTyRglnGluIleuIleThrPheIleGlyIleGlu 393
Db 1024 TTTCGCCGCCATTAATGAGGCCATGCTGCTGCTGCTGCTTATGTTCCGGGTATAGGC 1083
OY 394 ProAlaLys-----ThrProAlaGlnAlaGlyThrAspAlaLeuPro 407
Db 1084 CCGGTGACTCCGACACATCTGATTTACCCCAAGCTTGAAAGGCCAACAAGTGTCAAG 1143
OY 408 Ile-----LeuAlaLeuLysAspIleGlu-----AlaGluTyRysSerGlySerGlyLeu 424
Db 1144 TACTGGTAAAGCACGCGAGATGCCAATTGGCGAGGAAACATTTACGCGGTATGCTTGC 1203
OY 425 AsnValGlyTyRSerIleThrAsnThrGlyGlnGlnSerAsnLeuAspLeuGlyTyR 444
Db 1204 ACAACGGGAGACTGCTGTC-----1221
OY 445 ArgLeuAsnLysAlaAspGlyValIle-----GluValLysThrSerSer 460
Db 1222 -----GAGATTTTCCTCATATTGTTTCGAAAGAAACATGACAGCGAAG 1263
OY 461 IleAsnIleSerTrpTyRgIlyTyRgIlyGluHisProGlnSerPheSerLeuAlaProAs 480
Db 1264 CCGACCGCGCGATGAT-----GAAAGAACATCAACTTACCTGAAGGACTAAA 1314
OY 480 nglnLeuSerGlnGlyIleAsnThrIleThrLeu-----491
Db 1315 TACATCCGATGCGCGCATACACTGATCCGATATATATTCTTGAAGTTGACGATATC 1374
OY 492 -LeuTyRArgArgThrGlyThrGlnGlnTrpGluProValArgHisAlaGlnGlyTy 511
Db 1375 ACTGTATTTCGGG-ACCTCTGCATCAGAGCCGAACTGTT-----1413
OY 511 fValAsnSerIleLysValaAsnThrThraspProAsnAsnValValIleThrValAspAs 531
Db 1414 -----ACCGAT-----TTCCGTTCTCGCTTATTGAAAA 1442
OY 531 nasngIuglyLysLeuSer-----IleValProAsnSerPhe-----543
```

```
Db 1443 CAACAGGAGGAGATTAAGTCGAATTTCCTTAACGGCTCCGAAACCCGATTAAGCTGATGA 1502
OY 544 -----ValAlaAspLeuAsnSerTyRgIleHisSerThrIleThrValG1 558
Db 1503 TAAAGCCCATTCGAGCTCCCGCTACATATCTGCAAAAGCGCTGCTCTGTTCA 1562
OY 558 nPheAsnSerAspSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrG1 578
Db 1563 CATCAAGACCGC-----ACTGTTTGGAGTATCATCATGAGACTTA 1604
OY 578 yAlaThrAlaAspAspValIleSerLeu-----587
Db 1605 TTCTTCAGAGACGATCAGCATGGAAGTGAATATGTCTACCTGCCGTTTAAAGCAAA 1664
OY 587 -----587
Db 1665 TATCAGTCCCAATCGGTTTGGATTAAGCTGATTTATGATTCATCAATCGACATTAATCT 1724
OY 588 -----GlyTyRValMetAlaGluValProG1 596
Db 1725 ATATGAGGCTTTGAGCGCGGAGATATCTGAAAGCGTGCTGTGTTGATGCTGATGC 1784
OY 596 yGlySerSerAsnTyRProValValTyRSerLysAspValLeuThrLeuSerGlyLys 616
Db 1785 CGACAAATGTAAATGG-----GA 1802
OY 616 pTyRThrLeuTrpTyRArgPheSerIleAsnAsnGlnLysAspGluTrpLysIleG1 636
Db 1803 CTATTATCTGCTGACTATGATGGA-----CATGACAGGAGCAAGTATATGC 1850
OY 636 ySer-----ValSerValLysThrProThrGluTyR---ThnHisPr 649
Db 1851 ATCCCTTCCTACTTACCGATGATGCGCTTTTAACTCCGATTAACCTATTGTTGTTACAC 1910
OY 649 OleuPheGlnValGly-----HisAsnGlnThrSerThrTyRTh 662
Db 1911 CAACTTCGAAAGGAGCCAACTGTCAGATGATTTGGGTAAAGTGGCAAGATCGTGTATTC 1970
OY 662 rLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLysAsnGlyLeuPr 682
Db 1971 GGCTGAG-----CATTAATCGGTGATGCTTCTACTG-----GGAACTGC 2012
OY 682 oPheAsnGlyGlnLeuValValaPheArgGlnThrGlnSerSerGlySerLeuTr 702
Db 2013 TGTGAA---GATTTGCTCTCTGTTGGAAGACAAATGACCGCTTAAGCTAACGCTGC 2069
OY 702 pAlaAlaGlnGluThrValHisIleLysGlnGlyGluThrPhe-----716
Db 2070 ATGTATAGGAGCACTATTACATTTGCCGACGAAACAATAATATTTGCTGGCGGCATTA 2129
OY 717 -----ValTyRysProValVa 722
Db 2130 TGATTGCACCATATGTTTTTCTTCTGCTTGATGACATTAACGTTTATCTGTTTACGTA 2189
OY 722 IglnGlyProIleProAspGlySerTyRArgAlaThrLeu-----735
Db 2190 GACTGTTCCGAGCGCTGTACTGATTTCTGTGCTGCTGCTTATTAAGAAATACAAAGGTCG 2249
OY 735 -----735
Db 2250 CCTGAATGATGATTAATCTTAACGGCTACGACCAACCGATAAGACTGATTAATAAACCAT 2309
OY 736 -----HisAlaPheValaAsnGlyGlnGlnGlnLeuTyRLeuLys-----748
Db 2310 GCAGCTTACCGGCTACAAACATCTATGCAAAATGCTGCTCTCTTGTCAATACAAAGACC 2369
OY 748 -----748
Db 2370 GACTGTTTGGAGTATATCGATGAGACTTATCTTACGAGACGTCAGGTGGAATATGA 2429
OY 749 -----GlyLysAr 751
```

CC	be used to detect Porphyromonas gingivalis in standard hybridisation
CC	assays. Porphyromonas gingivalis is involved in periodontal disease
CC	especially gingivitis.
xx	
50	Sequence 2766 BP; 768 A; 580 C; 667 G; 751 T; 0 other;
	Alignment Scores:
	Pred. No.: 0.0015 Length: 2766
	Score: 159.50 Matches: 152
	Percent Similarity: 31.14% Conservative: 108
	Best Local Similarity: 18.20% Mismatches: 231
	Query Match: 3.59% Indels: 345
	DB: 20 Gaps: 38
US-10-030-330-1 (1-843) x AAX91739 (1-2766)	
QY	221 GlycerleuValGlyAsnTrpSerGlyThrPheGlyGluMetYrAspTrpIleAsnMet 240
DB	646 GGTGAGATGTGGAGCATGTGACTGTGAACTATCCGAGGATTAT-----CAA 693
QY	241 Pro-----GlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyr 257
DB	694 CCGGAGGAAAGGGGATGAAAGATTG---CAGTTAGCGGCTCAACATCTAGTGGAC 750
QY	258 AlaThrleuMetArgAspValSerAlaSerValSerMetSerPheTyrGluAsnGlySer 277
DB	751 GGTACACACTACAGGCACAAATGMAA-----GATGTCTCCATACAGTAGTATGTGAC 801
QY	278 GlyThrYrSerValYrValValGlyAlaLeuArgAsnPheAlaGlyTyrLysArgSer 297
DB	802 AGCACTTACTCT-----TTGCGAAGACAT-----CCC 828
QY	298 LeuGlnLeuHis-----ValArgAlaLeuTyrThrSerGlnGluTrpHisAspMetIle 315
DB	829 TTGCAAGTGGAGTACTGCGTTACAGCCGTTAC----- 861
QY	316 ArgGlyGluLeuAlaSerGlyArgProValTyrTyrAlaGluAsnAsnGlnSerIleGly 335
DB	862 -----GATGAAAGCATGAA 876
QY	336 HisAlaPheValCysAsp-----GlyTyrAlaSerAspGlyThrPheHisPheAsnTrp 353
DB	877 TCTTCGACCGATATGGGACGCTGCATTACCCACGAGATGCCATCTTATGAAATATTT 936
QY	354 GlyTrpGlyGlyValSerAsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeu 373
DB	937 GAGATGTGACTGTCTCCCAATGCTTG---CTT 966
QY	374 GlyIleGlyGlyGluGlyIleGlyPhe----- 382
DB	967 GTGATAGACGCTGATGGAGATGATTAAGTGGGACACTATTGTAATGCATACGACGCT 1026
QY	383 -----ThrIleTyrGlnGluIleIleThrGlyIleGlu 393
DB	1027 TTTCGCGGCATTAATGAGGCCATTGCTCTTGCGCTTATATGTTCGCGGATAGGC 1086
QY	394 ProAlaLys-----ThrProIleGluIleGlyThrAspAlaLeuPro 407
DB	1087 CCGGTGACTCCCGCAACACTATCTGATTATACCCCAAGGTGAGGACCCAAACGCTGTAA 1146
QY	408 Ile---LeuAlaLeuLysAspIleGlu-----AlaGluTyrLysSerGlnSerGlyLeu 424
DB	1147 TACGTGGGTAAAGCAGCAGGATGCCAATTTGGGCACGGAACTTACGGGGATAGCGCTTG 1206
QY	425 AsnValGlyTyrSerIleTyrAsnThrGlyGluGluGlnInsertAsnLeuAspLeuGlyTyr 444
DB	1207 ACAACGGGGACTGCTGTC----- 1224
QY	445 ArgLeuAsnLysAlaAspGlyGluValIle-----GluValLysThrSerSer 460
DB	1225 -----GGAGATTGTGCTATATTGTTCGACAGAACCATGACAGCGGAC 1266
QY	461 IleAsnIleSerTrpTyrGlyTyrGlyGlu---HisProGluSerPheSerIleuAlaProAs 480

[illegible]

PI Ross BC, Rothel LJ, Webb EA;
XX MPI: 1999-385613/32.
DR P-PSDB; AA134520.

PT Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis

PS Claim 12; Page 233-234; 588pp; English.

CC AA91536 to AA91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (Pg) polypeptide sequences given in AA934318 to
CC AA934583. AA931802 to AA931889 represent PCR primers used in the
CC isolation of the Pg polypeptides. The Pg polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The Pg polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gingivitis.

Sequence 2775 BP; 770 A; 582 C; 669 G; 754 T; 0 other;

Alignment Scores:

Pred. No.: 0.00151 Length: 2775
Score: 159.50 Matches: 152
Percent Similarity: 31.14% Conservative: 108
Best Local Similarity: 18.20% Mismatches: 231
Query Match: 3.59% Indels: 345
DB: Gaps: 38

US-10-030-330-1 (1-843) x AA91738 (1-2775)

QY 221 GlySerLeuValGlyAsnTrpSerGlyThrPheGlyGluMetTyraSptRileasmet 240
DB 655 GGTGAGATGTTGGGCGATGCTGCACTATCCGAGGATTAT-----CAA 702
QY 241 Pro-----GlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTy 257
DB 703 CCGGAGCAAGGGAATGAGAGCTG---CAGCTTACCGGCTACACATCATCGAAC 759
QY 258 AlaThrLeuMetAlaGAspValSerAlaSerValSerMetSerPheTyGluAsnGlySer 277
DB 760 GGTACACTACTGCGCAAAATGAAA-----GATGCTCCATAGCTAGATATGTGAC 810
QY 278 GlyThrTrpSerValTyValValGlyAlaLeuAlaGAsnAsnPhaGrtTyTyrArgSer 297
DB 811 AGCACTTCTCT-----TTGCGAGCAAT-----CCG 837
QY 298 LeuGlnLeuHis-----ValArgAlaLeuTyTrpSerGlnGluTrpHisAspMetile 315
DB 838 TTGCAAGTGGAGTACTGCTTACACGCGTTAC----- 870
QY 316 ArgGlyLeuLeuAlaSerGlyArgProValTyTrpAlaGlyAsnAsnGlnSerIleGly 335
DB 871 -----GATGAAAGCATAGAA 885
QY 336 HisAlaPheValCysAsp-----GlyTyraLaserAspGlyThrPheHisPheAsnTrp 353
DB 886 TCTTGACCGCTATGGGCGACCGCTGCATACGCGATGCGCATCTTATGAATAATTT 945
QY 354 GlyTrpGlyGlyValSerAsnGlyPheTyTrpLeuThrLeuLeuSerProThrSerLeu 373
DB 946 GAGATGAGACCTGTTCCCATGCTTGC-----CTT 975
QY 374 GlyIleGlyGlyGluGlyIleGlyPhe----- 382
DB 976 GTGATAGACGCTGATGAGATGATTTAGCTGGGACACATTTGAAATGATACACGCGT 1035
QY 383 -----ThrIleTyGlnGluIleIleThrGlyIleGlu 393
DB 1036 TTTCGGCGCATATGAGCCATTGCTTCTTGCGGCTTTTATGTTCCGGGTATAGCC 1095
QY 394 ProAlaLys-----ThrProAlaGluAlaGlyThrAspAlaLeuPro 407

DB 1096 CCGGTGACTCCGAGACAATATGATTAACCCCAAGCTTGAAGGACCAAGCGTGCAG 1155
QY 408 Ile---LeuAlaLeuLysAspIleGlu-----AlaGlyTrpLysSerGluSerGlyLeu 424
DB 1156 TACTGGGTAAAGCAGCGATGATGCAATTTGGGACCGGAGCAATTCAGCTGATGGCTCG 1215
QY 425 AsnValGlyTrpSerIleTyraSntThrGlyGluGlnSerAsnLeuAspLeuGlyTy 444
DB 1216 ACAACGGGACTGCTCTC----- 1233
QY 445 ArgLeuAsnLysAlaAspGlyLeuValIle-----GluValLysThrSerSer 460
DB 1234 -----GGAGATTGCGCATATGTTGGAAGAACAACATGACAGCGAG 1275
QY 461 IleAsnLysSerTrpTyGlyTyGlyGlu-HisProGluSerPheSerLeuAlaProAs 480
DB 1276 CCGACCGGCGCATGTTAT-----GAAGAACCATCACTTACCTGAGGACTTAA 1326
QY 480 nGlnLeuSerGlnGlyIleAsnThrIleThrLeu----- 491
DB 1327 TACATCGCATGGCGGCGATTAACACTGTACCGCATATATTTCTTGAAGTTGACGATATC 1386
QY 492 -LeuTyraArgArgThrGlyThrGluGlnTrpGluProValArgHisAlaGlnGlyTy 511
DB 1387 ACTGTATTTGGG-ACCTCGCATCAAGCCGCACTGTT----- 1425
QY 511 rValAsnSerIleLysValAsnThrThrAspProAsnValValAlaValAlaAspAs 531
DB 1426 -----ACCGAT-----TTGCTTCTCCCTTATGAAA 1454
QY 531 nAsnGluGlyLysLeuSer-----IleValProAsnSerPhe----- 543
DB 1455 CAACAAGGAGCATTTAACTGAATATATCTTAACGCTCAAGAACCCGATAGACTGATGA 1514
QY 544 -----ValAlaAspLeuAsnSerTyGluHisSerThrIleThrValG 558
DB 1515 TAAAGCCCATGTCAGCTTCCGCGTACAAATATCTATGCAAAACGCTCTCTTGTGA 1574
QY 558 nPheAsnSerAspSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrG 578
DB 1575 CATACAAAGCCCG-----ACTGTTTGGAGTATATGATGATGAGTCTTA 1616
QY 578 yAlaThrAlaAspAspValIleSerLeu----- 587
DB 1617 TTCTTCAGAGACATCAGCTGGAAGTGAATATGTGTCACTCCGTTATTAACGAAA 1676
QY 587 ----- 587
DB 1677 TATCGATGCCAATGCGTTTGGATAGCTGATATATGATTCATCGGACATTTATCTT 1736
QY 588 -----GlyTrpValMetAlaGluValProG 596
DB 1737 ATATGAAGCTTTGAGCCGGAAGTATTCCTGAAGGCTGTGTTGATGATGCTGATGG 1796
QY 596 yGlySerSerAsnTyProValValTrpSerLysAspValLeuThrLeuSerGluLys 616
DB 1797 CGACATGTAAATGG-----CA 1814
QY 616 pTyThrLeuTrpTyraArgPheSerIleAsnAsnGlnLysAspGluTrpLysLysIleG 636
DB 1815 CATATATCTTGGACTATATAGA-----CATGACAGTGAAGAAGTATTC 1862
QY 636 ySer-----ValSerValLysThrProThrGluTy---ThrHisPr 649
DB 1863 ATCCCTTCGTACTTACCGATGATGGCGTTTAACTCCGATTAATATTGGTTACAC 1922
QY 649 cLeuPheGluValGly-----HisAsnGlnThrSerThrTyTh 662
DB 1923 CAGACTCGAAGAGCAAGCTTGTCAAGTATTTGGTAAGTGGCAAGATGCTTTATATC 1982
QY 662 rLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLysAsnLeuGlyLeuPr 682


```

Db      925 TCTTCACCGGATGTCGACAGCTGCATTCACGCCACGAGCATCTTATGAAATTTT      984
Qy      354 G1YTPG1YGLYValSerAsnG1YPher1YrLysLeuLeuSerProThrSerLeu 373
Db      965 GAGAAATGACGCTGTTCCCAATGGTTGG-----CTT 1014
Qy      374 G1Y1LeG1YGLYGLY1LeG1YPh-----382
Db      1015 GTGATAGACGCTGATGAGATGATTTAGCTGGGACACTATTGATCATACACGCT 1074
Qy      383 -----Thr1YrG1NG1u1Le1YrG1Y1LeG1u 393
Db      1075 TTTCCCGGCATATGAGAGCCATTGCTCTGCGCTTCTATGTTCCGGTATAGGC 1134
Qy      394 ProAla1Ys-----ThrProAlaG1uAlaG1YThrAspAlaLeuPro 407
Db      1135 CCGGTGACTCCGACAACTATCTGATTCACCCCAAGGTGAAGGCCAAAGCTGTCAAG 1194
Qy      408 1Le---LeuAlaLeuLysAsp1LeG1u-----AlaG1u1YrLysSerG1uSerG1YLeu 424
Db      1195 TACTGGGTAGACACGACGATCCCAATTGGCGACGGACATTACCGGTGATGGCTTG 1254
Qy      425 AsnValG1YTYrSer1LeTYrAsnThrG1YGLYGLYGLYGLYGLYGLYGLY 444
Db      1255 ACAACGGGACGTGTCTC-----1272
Qy      445 ArgLeuAsnLysAlaAspG1YGLYGLYGLYGLY-----G1uVal1YrThrSerSer 460
Db      1273 -----GGAGATTTCGATATGTTGCGAAGAAACCATGACGAGAG 1314
Qy      461 1LeAsn1LeSerTYrTYrG1YGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLY 480
Db      1315 CCGACGGCGCATGGTAT-----GAAACACCATCACTTACCTGAAGGACTAA 1365
Qy      480 nG1uLeuSerG1NG1Y1LeAsnThr1LeThrLeu-----491
Db      1366 TACATCGCATGGCGCATTAACACTGATACCATATATATTCGTGAAGTTGACGATATC 1425
Qy      492 -LeuTYrArgArgThrG1YThrG1uG1uTrpG1uProValArgHisAlaG1NG1YGLY 511
Db      1426 ACTGATTTCGGG-ACCTCTGCATCAAGCCGACCTGTT-----1464
Qy      511 rValAsnSer1LeLysValAsnThrThrasProAsnAsnValValValThrValAspAs 531
Db      1465 -----ACCGAT-----TTCGTTGTCCTTATTTGAAA 1493
Qy      531 nasnG1uG1YLysLeuSer---1LeValProAsnSerPhe-----543
Db      1494 CAACAAAGGAGCATTAAGTGAATTATCTTAACGCTACGAACCCGATTAAGACTGATGA 1553
Qy      544 -----ValAlaAspLeuAsnSerTYrG1uHisSerThr1LeThrValAl 558
Db      1554 TAAAGACCCATGTCAGCTTCGCGGTACATATCTGCAAAAGCGCTCCCTCTGTTCA 1613
Qy      558 nPheAsnSerAspSerProAspG1u1LeArgThrProValAlaPheAlaLeuSerThrG1 578
Db      1614 CATCAACAGCCG-----ACTGTTTGGAGTATATGATGAGACTTA 1655
Qy      578 yAlaThrAlaAspAspVal1LeSerLeu-----587
Db      1656 TTCCTCACGACGACATCAAGTGAAGTGAATATTGTCTACCTGCCGTTTATACAGCAA 1715
Qy      587 -----587
Db      1716 TATCAGTCCCAATGCGTTTGGATAGCTGATTTATGATTCATCGACATATATCTT 1775
Qy      588 -----G1YrPValMetAlaG1uValProG1 596
Db      1776 ATATGAAGCTTTGAGCGCGAAGATATCTTGAAGGCTGTGTTGATGATGCTGATG 1835
Qy      596 yG1YSerSerAsnTYrProValVal1YrPserLysAspValLeuThrLeuSerG1uG1Y 616

```

```

Db      1836 CGACAACTGTAATTCG-----GA 1853
Qy      616 pTYrThrLeuTrpTYrArgPheSer1LeAsnAsnG1uLysAspG1uTrpLys1LeG1 636
Db      1854 CTATTATCTTGGACTATGATGA-----CATGACAGTGAAGTGTATTCG 1901
Qy      636 ySer-----ValSerVal1YrThrProThrG1uTYr---ThrHisPr 649
Db      1902 ATCCCTTCGACTTACCGATGATTTGGCGTTTAACTCCGGATTAACATTTTGGTTACCC 1961
Qy      649 oLeuPheG1uValG1Y-----HisAsnG1uThrSerThrTYrTh 662
Db      1962 CAGACTCGAAGAGACCAAGCTGTCGAATGATGGTAAGTGGCCAGATGCTGTTATTC 2021
Qy      662 rLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLysAsnLeuG1YLeuPr 682
Db      2022 GCGTGA-----CATATGCTGTGATGCTTCTACTACG-----GGAACGTC 2063
Qy      682 oPheAsnG1YGLYLeuValValAlaPheArgG1uThrG1uSerSerSerG1YSerLeuTr 702
Db      2064 TGTGAA-----GATTTGCTCTCTTTCGAAAGACAAATACCGCTTAAGCTAACGTCG 2120
Qy      702 pAlaAlaG1uG1uThrValHis1LeLysG1NG1YGLYGLYGLYGLYGLYGLY 716
Db      2121 ATGCTATGAGCGCACTATATACATTCGCTGCAGAACAAATATTCCTGCGGCATTA 2180
Qy      717 -----ValTYrLysProValVa 722
Db      2181 TGATTGACCCGATATGTTTCTTCTGCTTGGATGACATTAACGTTATGCTTACTGA 2240
Qy      722 1G1uG1YPro1LeProAspG1YSerTYrArgAlaThrPhe-----735
Db      2241 GACTGTCCGAGCCTGTTACTGATTTTCGTTCTCGCTTATGAGATAACAAGGCTCG 2300
Qy      735 -----735
Db      2301 CCGTAATGGAATTAATCTTAACGCTACGAACCCGATTAAGCTGATTAATAACCATT 2360
Qy      736 -----HisAlaPheValAsnG1YGLYGLYGLYGLYGLYGLYGLYGLY 748
Db      2361 GCAGCTTACCGGCTACACATCTATGCAATAGGCTCGCTCTGTTTCATACAAAGACCC 2420
Qy      748 -----748
Db      2421 GACTGTTTGGACTATATGATGATGACTTATTTCTTACGAGAGCTGAGTGAANTGA 2480
Qy      749 -----G1YLysAr 751
Db      2481 ATATTGTGTCACGCGCTTATTAACGAACAATATCGAGTCCCATCGTTTGCATTAAGCT 2540
Qy      751 gAsnTYrThrValLys1LeValAsnG1YThrAlaValAlaG1u1LeG1uSerSerG1uG1 771
Db      2541 GAACATACTATC-----ACATCTTGGATTAATATTCATCACTATACAG 2585
Qy      771 u1LeArgValPheProAsnProAlaArgAspTYrValG1u1LeSerAlaProCys1LePr 791
Db      2586 CTTGAATAATATATCCATACCGCATCGATGCGTATGAGATGAGAGGA---TTGAGTCG 2642
Qy      791 oG1uG1uThrSer1Le1LeuPheAspLeuSerG1YLys1LeValMetLysAsnSerLe 811
Db      2643 GAGCAAGTCGACATGAGATGATTAATGCGGTGGAATTTGATTAAGGAGAAAGAC 2702
Qy      811 uSerAlaG1YHisG1YArgMetAspValSerArgLeuProAsnG1YAlaTYr1LeLeuY 831
Db      2703 TCATTGAGAAAGAAAGCAATGATGTTTCACCTCAATGACGAGGAGTCTACTTGAATTA 2762
Qy      831 sValAspG1Y-----TYrThrTYrLys1LeAsn1Le 841
Db      2763 AGTAGTCGTTGGAATTAATAACAACACCGAAAGGTAGAGATA 2805

```

RESULT 17
 AAH81312
 ID AAH81312 standard; DNA; 3978 BP.

XX AAH81312;
AC
XX
DT 21-SEP-2001 (first entry)
XX
DE Escherichia coli protein encoding nucleotide sequence SEQ ID NO:111.
XX
XX Escherichia coli; identification; proliferation; microorganism;
KM antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
XX bacterial growth inhibition; de.
XX
OS Escherichia coli.
XX
XX WO200148209-A2.
XX
PD 05-JUL-2001.
XX
PF 19-DEC-2000; 2000WO-US34419.
XX
23-DEC-1999; 9905-0173005.
(ELIT-) ELITRA PHARM INC.
XX
PI Forsyth RA, Ohlsen KL, Zyskind JW;
XX
DR WPI: 2001-457376/49.
XX
P-PSDB: AAG98256.
XX
PT Novel nucleic acids encoding proteins required for Escherichia coli
XX proliferation, useful for screening for antimicrobial agents -
PS
XX Claim 9; Page 148-154; 596pp; English.
XX
XX The present invention describes a purified or isolated nucleic acid
CC sequence (1) consisting essentially of one of the 93 nucleotide sequences
CC given in AAH81202 to AAH81294, where expression of the nucleic acid in a
CC microorganism is capable of inhibiting proliferation of a microorganism.
CC (1) have antibacterial and antibiotic activities, and can be used in
CC gene therapy. Expression of (1) in a microorganism inhibits proliferation
CC of the microorganism, and the manufactured antibiotic is useful for
CC reducing the activity or level of a gene product required for
CC proliferation of a microorganism in a subject, specifically humans. The
CC nucleic acids that inhibit bacterial growth or proliferation can be used
CC as antisense therapeutics for killing bacteria. In addition to
CC therapeutic applications, the nucleic acid sequences complementary to
CC sequences required for proliferation can be used as diagnostic tools.
CC For example, nucleic acid probes complementary to proliferation-regulated
CC sequences that are specific for particular species of microorganisms can
CC be used as probes to identify particular microorganism species in
CC clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli
CC proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491
CC represent oligonucleotides, which are used in the exemplification of the
CC present invention.
XX
XX
SQ Sequence 3978 BP; 1071 A; 819 C; 1143 G; 945 T; 0 other;
Alignment Scores:
Pred. No.: 0.00715 Length: 3978
Score: 154.00 Matches: 159
Percent Similarity: 33.97% Conservative: 89
Best Local Similarity: 21.78% Mismatches: 252
Query Match: 3.47% Indels: 230
DB: 22 Gaps: 40
US-10-030-330-1 (1-843) x AAH81312 (1-3978)
OY 161 ASnglYhIsPhaAlaAsnAspPromeTatgTTPasnglGlyTYrProTTPasnsnlsy 180
DB 883 AATGATATTCGATATGACAGTCATATTTCCATATATGACTGATATATACAA 942
OY 181 GluProlLeu-----ProAsnglYhAsnHIsAlaTYrThClYcysValAlaThr 197
DB 943 GAATATTTCATTGCGGCGCTTCAAGGAGGCTCCACGGGTGCTCAACGTGACCGATAA 1002

OY 198 AlaAlaAlaGlnIleMetArgTYrHISerTyrProlLeuGlnGlyGluGlySerPheAsp 217
DB 1002 ----- 1002
OY 218 TYrHISAlaGlySerLeuValGlyAsnTPSer-----GlyThPheGlyGlu----- 233
DB 1003 ----- 1003
OY 234 MetTYrAspTPRILeAsnMetProGlyAsnProAspLeuAsnLeuThrGlnSerGln 253
DB 1039 CCCTATATCTATATGCTGATGCTGCGGCGGATTAATGCTTCGATGAGGCAAA 1098
OY 254 ValAspAlaTYrAlaThrLeuMetArgAspValSerAlaSerValSerMetSerPheTYr 273
DB 1099 GTAGAT-----TCGGAAATTATCATCTCGGGGATGAA 1131
OY 274 GluAsnGlySerGlyThTYrISerVal-----TYrValAlaGlyAlaLeuArg 289
DB 1132 GAACAGGCGACAGGCAACATTAATGATTAAGATTAAGAACTCCGTATCATCTATCGGA 1191
OY 290 AsnAsnPheArgTYrLYsArgSerLeuGlnLeuHISVal-----ArgAlaLeuTYrThr 307
DB 1192 ACTAATCTTGTTATGACGCCACAGCGGCAAAATGAAATATCATGATGAGGCTGTGTTC 1251
OY 308 SerGlnGluTYrPHisAspMetIleArgGlyGluLeuAlaSerGly----- 322
DB 1252 AGCAACGAGGAAAGTTCACTCGGTATGAGAAACCGCGCGGAAATGACATCACC 1311
OY 323 -----ArgProValTYrTYrAlaGlyAsnAsnInsSerIle 334
DB 1312 ACGGGGGAATGTGGAGGTCAATGAATGCTATACACCATTCGTTGCGGCGGC 1371
OY 335 GlyHISAlaPheValCysAspGly-----TYrAlaSerAspGlyThrPheHISphe 351
DB 1372 GGAAACCTTAATATTAATGATGCGGATGCGGTATGCGCAAAATATTAATCTTT----- 1425
OY 352 AsnTPRgLYrTYrGlyValSerAsnGlyPheTYrLYsLeuThrLeuLeuSerProThr 371
DB 1426 -----TTGGGCGATTAACAAACGCGTATGCGACACTGATGATGATGACGA 1476
OY 372 Ser-----LeuGlyIleGlyGlyGlyIle----- 380
DB 1477 TCATGCTTCGAACTGTGGGATTCATATGCTGAATTTGTAGGCGGATTCATAATGTC 1536
OY 381 -----GlyPheThrIleTYrGlnGlu-----IleIleThrGlyIleGluProAla 395
DB 1537 AGTATGTCGCCACCTTAATTCACAGGCGCTATGATTTATGCGGAAATATGCTCCGCT 1596
OY 396 LYsThrProAlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLeuAspIleGlu 415
DB 1597 AAGGGATAGTAAATTAATTTCAACGCGACACTCTCGAAATTTAAGCATCATCTACTAAC 1656
OY 416 AlaGluTYrLYsSerGluSerGlyLeuAsnValGlyTYrSerIleTYrAsnThrGlyGlu 435
DB 1657 GCACAAATTC-----CTACAGTCTCGT-----GTATTAACGACCGGGTGAA 1695
OY 436 GluGlnSerAsnLeuAspLeuGlyTYrArgLeuAsnLYsAlaAspGlyGluValIleGlu 455
DB 1696 -----CTGAATATTAACACCGGAGCTATAGTATAA 1725
OY 456 ValLYsThrSerSerIle-----AsnIleSerTYrGly 467
DB 1726 GCGCGTATGACACAGTATGCTCTCAATGACAAAGTAAAGCGAGCTGAGGTGATGGG 1785
OY 468 TYrGlyGlnHISProGluSerPheSerLeuAlaProAsnGlnLeuSerGlnGlyIleAsn 487
DB 1786 CAGAACTCTCTCTGTAAGCAATATGATGATGATGATGATGATGATGATGATGATGAT 1839
OY 488 ThrIleThrLeuLeuTYrArgArgTYr-----GlyThGlnGluInTP----- 501
DB 1840 ACCTTAACCTTACGAAATACGATGATGATGATGATGATGATGATGATGATGATGATGAT 1899

```

OY 502 -----GluProval-----ArgHisala 507
DB 1900 GTTTTGGACCTGCTGAGCAAGCTAAACATTGGTGTCTGCTACAGGTGAGCGGCGCA 1959
OY 508 GlnGlyGlyTyrVal---AsnSerIleLeuVal----- 517
DB 1960 GATCGCGGATTATATACCAATGCGACAGAAAGTGAGTTTGTCTTGGCGAAGCGCTTTT 2019
OY 518 -----AsnThrThrAspProasn-----AsnValValIleThrVal 529
DB 2020 GTCTTTAATCATACCAATACAGTATGCGGCTACAGGTGCGATGCTGATTACAGGT 2079
OY 530 AspAsnAsnGlnGlyLysLeuSerIleValProAsnSerPheValAlaAspLeuAsnSer 549
DB 2080 GACGATTAAGACGGAAAGTATC----- 2103
OY 550 TyrGlnHisSerThrIleThrValGlnPheAsnSerAspSerProAspGlnIleArgThr 569
DB 2104 -----CATGATGACGAGCCATACGGTGTTCATGACGAGGAATACTTATACGGTAAACG 2157
OY 570 ProValAlaPheAlaLeuSerThrGlyAla-----ThrAlaAspValIleSerLeu 587
DB 2158 CTGGTCAATGACGGCTCTCTGATCCATGCGCTCATGCGGAGATGGGTTACGGGCAAG 2217
OY 588 GlyTrpValMetAlaGluVal-----ProGlyGlySerSerAsnTyrProVal 603
DB 2218 GGG-----TCGAGTGAAGTAAACCATTCGCAACCCCGT----- 2250
OY 604 ValTrpSerLysAspValLeu-----ThrLeuSerGlnGlyAspTyrThrLeuTyrPyr 621
DB 2251 -----ACGTCGACATTCGTCATCAACGACAGTGCAGAGATACAGCTGACCAAT 2304
OY 622 ArgPheSer-----IleAsnAsnGlnLysAspGluTrpLysIleGlySer 637
DB 2305 CGCGTCAAGCGGATGCTGATGCGAGTGCAGTGCATCTCCGACAAAGATGTTTGGC 2364
OY 638 ValSerValLysThrProThrGluTyrThrHisProLeuPheGluValGlyHisAsnGln 657
DB 2365 TTTACCCATGCAACGAGGACTGAAATTCGCC-----GGTTTGCCTCACTGAAA 2412
OY 658 ThrSerThrTyrThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLys 677
DB 2413 GACAGTACCTTCACCTCTGGAACGCGACAACCGCTGCGCTTACTACGCGCATGTTGCAG 2472
OY 678 AsnLeuGlyLeuProPheAsnGlnGlyLeuValValPheArgGlnThrGlnSerSer 697
DB 2473 TCT-----GACAGTGAATAATACCATCGGTAAGAGTTGAGAGCAATCC 2517
OY 698 SerGlySerLeuTrpAlaAlaGlnGlnIleThrValHisIleLysGlnGlyGluThrPheVal 717
DB 2518 ATTGGTGCATG-----GCCATGCAATGGAGGTACCATCATT 2553
OY 718 Tyr-----LysProValValGlnGlyProIleProAspLysSerTyr 731
DB 2554 TTCGATACGATATATCTGCTGCGACGCTCGCGGAGGATATATC----- 2598
OY 732 ArgAlaThrLeuHisAlaPheValAsnGlnGlnGlnGlnLeuTyrLeuLysGlyLysArg 751
DB 2599 -----ACGTCGATACCGCTGGTGTGCGCGCGGCTACTACCTCGGAAAAGGC---CGT 2649
OY 752 AsnTyrThrValLysIleValAsnGlnTyrThrAlaValGluAlaIleGluSerSerGlnGlu 771
DB 2650 AACTATACG-----GTAACGGAGACGGCGCGACGTCCTATATGAC----- 2688
OY 772 IleArgValPheProAsnProAlaArgAsp 781
DB 2689 -----GTGCCCTAAACCGTGAATGAT 2709

```

RESULT 18

AAH75059

ID AAH75059 standard; DNA; 4059 BP.

AC AAH75059;

```

XX 29-OCT-2001 (first entry)
DT Nucleotide sequence of xyloglucanase.
XX xyloglucanase; family 44; glycosyl hydrolase; detergent;
DE cellulosic fiber; textile scouring; ss.
XX Paenibacillus polymyxa.
OS
XX Key Location/Qualifiers
XX CDS 1..4059
XX FT /tag= a
XX FT /product= "xyloglucanase"
XX
XX WO200162903-A1.
XX
XX 30-AUG-2001.
XX
XX 21-FEB-2001; 2001WO-DK00116.
XX
XX 24-FEB-2000; 2000DK-0000291.
XX
XX (NOVO ) NOVOZYMES AS.
XX
XX Schnorr K, Jorgensen PL, Schuelein M;
XX WPI; 2001-522819/57.
XX P-PSDB; MAG63962.
XX
XX New xyloglucanase enzyme belonging to glycosyl hydrolases family,
XX useful for detergent compositions, and textile or cellulose fiber
XX processing industries
XX
XX Claim 14; Page 80-81; 97pp; English.
XX
XX The present sequence encodes a xyloglucanase of the invention. The
XX specification describes a xyloglucanase enzyme belonging to family 44
XX of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
XX of at least 30% at pH 5-8. The enzyme exhibits high performance in
XX detergent compositions and prevents binding of certain soils to the
XX xyloglucan left in the cellulosic material. It is stable at pH 5-10 at
XX room temperature and has a half life of more than 50 days when incubated
XX in a full formulated liquid detergent at 30 plus degrees Celsius. The
XX enzyme is used in detergent compositions, textile industry for improving
XX the properties of cellulosic fibers, yarn, and woven or non-woven
XX fabrics, preferably in textile scouring process, and in cellulose fiber
XX processing industry for ratting of fibers e.g. hemp, jute, flax, and
XX linen.
XX
XX Sequence 4059 BP; 1187 A; 893 C; 1046 G; 933 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 0.0214 Length: 4059
XX Score: 148.50 Matches: 176
XX Percent Similarity: 32.10% Conservative: 127
XX Best Local Similarity: 18.64% Mismatches: 334
XX Query Match: 3.35% Indels: 307
XX DB: 22 Gaps: 47
XX
XX US-10-030-330-1 (1-843) x AAH75059 (1-4059)
XX
XX 112 AspMetAspSerMetProAspAsnLeuAlaGlyMetTrpLeuGlnIleTyrAspGlnIle 131
DB 844 GATTGGGACTCTGTAAGCAATATATACCTGGTGTGAGACTATATACCTGATCAAAAG 903
OY 132 GlyLeu----- 133
DB 904 CGCCTTAGCTCGCAAGTCAAGCAAGACATTCCTGATCTATTCACGCTACACTGATAT 963
OY 134 -----IleLeuSerGlyLysAlaGlnLeuAsnGlnIleLeuArgThrGlnGlyVal 151
DB 964 CCCGAAGCATGGCGGAGGACATACGAATATACGAATGAGGTAGCGAATACGAAGCAAG 1023

```

```

OY 152 ProAlaGluValIHisAlaLeuMetaspasngLHisPheAlaAsnAspPrometArgTyr 171
Db ||| ||| |||
1024 AAGCGCAATGACAGCA-----CCTGCACACTTGTGG 1056
OY 172 AsnGlnGlyTyr-----ProTyrAsnAsnLys-----GluProLeu 183
Db ||| ||| |||
1057 GACCGCACTTAAGAAGATAGTGGATGCTCATGCAACAGCAGATTTTGGCCATA 1116
OY 184 LeuPro-----
Db ||| ||| |||
1117 CTACCTCGATTGAAGCAGTGGTGCAATAATATTATCCGGAACCAAGCTGCATGACC 1176
OY 186 -----AsnGlyAsnHisAlaLysTyrHisGlyCysValAlaIleAlaAla---- 200
Db ||| ||| |||
1177 GATTATAGTATGCGCGCAAAATGATATTTCGGCGGGATTCGATACCGATGTGCTG 1236
OY 201 -----GlnIleMetArgTyrHisSerTyrProLeuGln-----GlyGlu 213
Db ||| ||| |||
1237 GGTATCTTGGGCAAAATGATGATTATATGCGCAAACTACTGGAAGCTAAAGATGTGTC 1296
OY 214 GlySerPheAspTyrHisAlaGlySerLeuValGlyAsnTyrSerGly-----Thr 230
Db ||| ||| |||
1297 AACCACTACGTTAGTCCGCTTACCAAGCTTATATGCAATATGCGAATAAACTCTACT 1356
OY 231 PheGly-----GluMetTyrAspTyrPheAsnMetProGlyAsnPro 244
Db ||| ||| |||
1357 TTCGCTGATACCAAGTGTATGCGCAACATCGGATTTGTCTCAATAGCTCGTCATGCT 1416
OY 245 AspLeuAspAsnLeuThrHisGlnSerGlnValAlaAspAlaTyrAlaThrLeuMetArgAspVal 264
Db ||| ||| |||
1417 TCTGTACGAATGATCGACGCAAAAGAACTGCAT---CTCGTGTCTCATGAATAAAGCATG 1473
OY 265 SerIleAspValSerMetSerPheTyrGluAsnGlySerGlyTyrHisTyr---SerValTyr 283
Db ||| ||| |||
1474 GACGCGCACTTCACGCGCCCAATTGATCTTCGCGCGCAAGACTTACATTTCCGCTAAA 1533
OY 284 ValIleGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGlnLeuHisVal--- 302
Db ||| ||| |||
1534 GTATGGGGTTCATAAACAGCTGCGCAAAATTAAGAAGCAGCCCAATCACCCAATT 1593
OY 303 -----ArgAlaLeuTyrThrSerGlnGluTyrHisAspMetIleArgGlyLeu 319
Db ||| ||| |||
1594 TCAGCAACCGTTTACTTATACCTGACCG-----CCTTGG 1629
OY 320 AlaSerGlyArgProValTyrTyrAlaGlyAsnGlnSerIle-----GlyHis 336
Db ||| ||| |||
1630 ACGCATATACATGTTGCTGCTACTGCTAGCTGCAATGACAGCTCTCCAGGAAAGCTCGAA 1689
OY 337 AlaPheValCysAspGlyTyrAlaSerAspGlyTyrPheHisPheAsnTyr----- 353
Db ||| ||| |||
1690 AGCTTTAAGCTGAAGCTGAGGCTGATGATGGGAAAGTCAATTATCTGGATGCTTCC 1749
OY 354 -----GlyTyrGly----- 356
Db ||| ||| |||
1750 AGCGAGTGTAGATACAGCGTACAGCGGCAACAGATGAACCGCCCTTCACTGCT 1809
OY 357 -----GlyValSerAsnGly----- 361
Db ||| ||| |||
1810 GTAGCATCCAACTTGACCGCAACGCTTATACGATACATCAACAGCTACAAAGCTTCA 1869
OY 362 -----PheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGluGly 1930
Db ||| ||| |||
1870 TACTATTACAAAGTAC---GCCAATACCAATTAAGGATGAGGATGCAATCAATAT 1923
OY 381 GlyPheThrIleTyrGlnGlnIleIleThrGlyIleGluProAlaLysThrProAlaGlu 400
Db ||| ||| |||
1924 TGAAGAGCGGTTCCGAAGATGCTGTAAACGT-----CCGCTGCTATGAAGCCGAA 1977
OY 401 AlaGlyTyrAspAlaLeuProIleLeuAlaLeuLysAsp-----IleGluAlaGluTyr 418
Db ||| ||| |||
1978 GAAGCAGC-----CTGAAGGAACCATTTGGAATCAGCGGG 2016

```

```

OY 419 LysSerGluSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGlyGlnGlnSer 438
Db ||| ||| |||
2017 ACCGCTACTCCGGTCTGCTGTTATGACGAATTCACAAATCCACAGGATCTCTGAC 2076
OY 439 -----AsnLeuAspLeuGlyTyrArg----- 445
Db ||| ||| |||
2077 ATGACATTCAGCTCCACAGCGAGCTTGTCATCAATCTTACAAATGCGCTACCTCTCT 2136
OY 446 -----LeuAsn-----LysAlaAspGlyGluValIleGlu 455
Db ||| ||| |||
2137 CATGATGACAAACGCCCAATTTCTCATTAACGCGCAAAAGGTTTGGGAAGCTCTCT 2196
OY 456 ValLysThrSerSerIleAsnIleSerTyrGlyTyrGlyGlnHisProGluSerPhe 475
Db ||| ||| |||
2197 AAGAAACCGCTGATTT----- 475
OY 476 SerLeuAlaProAsnGlnLeuSerGlnGlyLysAsnThrIleThrLeuTyrArgArg 495
Db ||| ||| |||
2224 TCCGAGAGCAAGCTGCTGTGTGAATGACAGCGCGCAATACGATCAGTTT-----GAA 2274
OY 496 ThrGlyThrGlnGlnTyrProValArgHisAlaGlnGlyLysTyrValAsnSerIle 515
Db ||| ||| |||
2275 ACAGCTGG---GGCTGCTACGATATC-----GACTACGTCAGACTGAG 2316
OY 516 LysValAsnThrThrAspProAsnAsnValValAlaThrValAspAsnAsnGlnGlyLys 535
Db ||| ||| |||
2317 CCTCGCCGTCAGCGCCCACTCATGCGGTAAACCAAAACGCTTACCAAT----- 2364
OY 536 LeuSerIleValIleProAsnSerPheValAlaAsp-----LeuAsn 548
Db ||| ||| |||
2365 -----CCGATGCGAGCGGTAGAGCAAAAGCAATTCATGATCTGCTGAT 2412
OY 549 SerTyrGluHisSerThrIleThrValGlnPheAsnSerAspSerProAspGluIleArg 568
Db ||| ||| |||
2413 CAATACGGAAGAATGCTCTCTGTCAGAGAAATTAACGAATATGATGCTTCA 2472
OY 569 ThrProValAlaPheAlaLeuSerThrGlyAlaThrIleAlaAspAspValIleSerLeuGly 588
Db ||| ||| |||
2473 GCCATGTA-----GTAATAAGCGGCGGATTCGAGCGCTGACCTGATGCACTATTCG 2526
OY 589 TrpValMetAlaGlu-----ValProGlyLysSerSerAsnTyrProValIleTyrSer 606
Db ||| ||| |||
2527 CCAGCAGAGCGGAACAGCTTATGTTCCACAGAGCAAGCAAAAGCGATTCAGAT 2586
OY 607 LysAsp-----ValLeuThrLeuSerGlnGlyAspTyrThrLeuTyrPhe 621
Db ||| ||| |||
2587 AAGCAAGGGGATCGTTTACCTTTGCA-----TGCACAGCAAGCA 2628
OY 622 ---ArgPheSerIleAsnAsnGln---LysAspGluTyrLysIleGlySerValSer 639
Db ||| ||| |||
2629 CCGAAGGTCTGATTCGATACGAGGGAAGAAATGCTGGAAGGCTTATGCTCATTC 2688
OY 640 ValLysThrProThrGlnTyrThrHisProLeuPheGluValGlyHisAsnGlnThrSer 659
Db ||| ||| |||
2689 ACCACATTCATTAATATGCG-----ATGAATCATCCAGAGTCCGA 2733
OY 660 ThrTyrThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLysAsnLeu 679
Db ||| ||| |||
2734 GATTATTAATTTACTTATTCGACATCATGATGATGAGG---CAATTGAACATTTG 2790
OY 680 GlyLeuProPheAsnGlyGluLeuValValPheArgGlnThrGlnSerSerGly 699
Db ||| ||| |||
2791 CAG-----GATGCGAAGGTTCTGCTCTGCTCTGCTTGCAGCAAGCGGAAGCA 2841
OY 700 Ser-----LeuTyrAlaIleGlnGlnIleThrValHisIleLysGln----- 712
Db ||| ||| |||
2842 AAATGTTCTGTGGGCGCAAGGCTCGAGGCTGTAAAGCTGTATATTTAATG 2901
OY 712 ----- 712
Db ||| ||| |||
2902 CACGACGTTTGAGGATGTCACAAATTCGAATTTGGGTATGAAATCTGTT 2961
OY 713 -----GlyGluThrPheVal----- 717

```

```

Db 2962 GCTCCGATGGTATCGGAGACAGATATGATATTTGAGCTTGACTTATCCG 3021
Qy 718 -----TyrLeuProValVal----- 722
Db 3022 CAGCAGGTGATACAGCCCGCAAAATTCAAAATACGAAACCTTGTTCATTTGGCGAAG 3081
Qy 723 -----GlyLeuProValProVal-----Gly 729
Db 3082 GACAAAAGCTAGTTGCCATGACGCAAAATGAGCGATCCGAGCCCTGATTTGATGAAG 3141
Qy 730 SerTyrArgAlaThrLeuHisAlaPheValAsnGlyGlnGlnLeuTyrLeuGlyGly 749
Db 3142 GCGTATCAAGCTCATTTGAGAGTGGTTCGCTACATGATGATGAGATTTTGTGAGAGCGGC 3201
Qy 750 LysAlaGln-----TyrThrValLysIleValAsnGlyThrAlaValAlaIle 766
Db 3202 AAACAAAAGCCTTGACCATCTGAAAAAGGTATATCATCCGAAACCTCATTAAGCTG 3261
Qy 767 GluSerSerGluGluIleArgValPheProAsnProAlaArgAspTyrValGluIleSer 786
Db 3262 GATGAG-----CTCCACAGCAACTTAAAAACGATATGCGATTACTGAG 3303
Qy 787 AlaProCysIleProGlnGluThrSerIleIleLeuPheAspLeuSerGlyLysIleVal 806
Db 3304 CAGCCGTCGTCACGCGGAGCTTCACGCTGAACGCTGCGGGTGAAGAGCGCAAGATACG 3363
Qy 807 MetLysAsnSerLeuSerAlaGlyHisGlyArgMetAspValSerArg-----LeuPro 824
Db 3364 CTAACTGACAGCATCTGCGCGCAATGCGAAACCTATGAAGTGAAGCTTCGACGACTGAA 3423
Qy 825 AsnGlyAlaTyr 828
Db 3424 AACGCGCGCTTC 3435

RESULT 19
AA12949
ID AA12949 standard; DNA; 9047 BP.
XX
AC AA12949;
DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:12.
XX
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
vaccine; attenuation; computer readable medium; ds.
Enterococcus faecalis.
MO9850555-A2.
12-NOV-1998.
PF 04-MAY-1998; 98WO-US08985.
PR 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046655.
XX
PA (HUMA-) HUMAN GENOME SCI. INC.
PI Barash SC, Dillon PJ, Kunsch CA.
XX
DR WPI: 1999-045171/04.
XX
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
infection.
XX
PS Claim 1; Page 305-310; 2084pp; English.
XX

```

```

CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AA12938 to AA13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
SQ Sequence 9047 BP; 3107 A; 1542 C; 1841 G; 2551 T; 6 other;

Alignment Scores:
Pred. No.: 0.0631 Length: 9047
Score: 148.50 Matches: 182
Percent Similarity: 34.59% Conservative: 120
Best Local Similarity: 20.85% Mismatches: 320
Query Match: 3.35% Indels: 252
DB: Gaps: 44

US-10-030-330-1 (1-843) x AA12949 (1-9047)
Qy 143 GluGluIleLeuArgThrGluGlyValProAlaGluValHisAlaLeuMetAspAsnGly 162
Db 1074 GAAAAAGTAAATCAGACGAGGATTTTCAGCAGCGCAC---CATTTGTGCGG----- 1121
Qy 163 HisPheAlaAsnAspProMetArgTyrPasnGlnGlyTyrProTyrPasnAsnLysGluPro 182
Db 1122 -----AATGCCCAAGTACTGTAATAC---CAATTG 1151
Qy 183 LeuLeuProAsnGlyAsnHisAlaTyrThrGlyCysValAlaThrAlaAlaGlnIle 202
Db 1152 TTGCATCCAAAGGCAAT-----GGACAGCGGCGCAAAAGATT 1190
Qy 203 ---MetArgTyrHisSerTyrProLeuGlnGlyGlySerPheAspTyrHisAlaGly 221
Db 1191 GTCATTTGATTTTATGATACGAATGACAAAGGAAACCATTTTGAACACACGCGCT 1250
Qy 222 SerLeuVal-----GlyAsnTrpSer---GlyThrPhe----- 231
Db 1251 CCGATTATTAACGGTGGTGGCACCTACGACAGGAGCTTCAACGTCGATCTAGGG 1310
Qy 232 -----GlyGluMetTyrAspTyrPrlLeuAsnMet-----ProGlyAsn 243
Db 1311 GCGGTGCACTTAGTGAACCAAGATATTTGGATATTTCAAAATTTAGACGACAAATACA 1370
Qy 244 ProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThrLeuMetArgAsp 263
Db 1371 CCAGACCTAGATATATCTC-----GAAGCGCATAGAAAGAACCCAGAGATGCT 1415
Qy 264 ValSerAlaSerValSerMetSerPheTyrGluAsnGlySerLysThrTyrSerValTyr 283
Db 1416 CAAGAGCGCGGATATTAAGTACTAGATATGAAACAAATGCAAGCTTCATTAAGTCTC--- 1472
Qy 284 ValValGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGlnLeuHisValArg 303
Db 1473 -----ACGATTAAGAAATACATGTTCAATGAT----- 1499
Qy 304 AlaLeuTyrThrSerGlnGluTyrPrlHisAspMetIleArgGlyLeuAlaSerGlyArg 323
Db 1500 -----GTTCAACAGCAATATTTATTAATTAAGTGAAGTCTGCAACCAACGA 1550
Qy 324 ProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCysArgGlyTyr 343
Db 1551 -----TTAAAGCGGATGAGTGAATATTTGTTTAGAGACTGG 1589
Qy 344 AlaSerAspGlyThrPheHisPheAsnTyrGlyTyrGlyVal----- 358

```

Db 1590 TTGTATGAGAAAT-----GGCAATGTAGTAACGGCAGCTAAT 1625
 Oy 359 -----SeranSglyPheTyrLysLeuThrLeuLeuSerProthrSerLeuGly 374
 Db 1626 GATCATCCGACACACGCTGATTAATGATATTGTTGAAATTAATGATTACACA 1685
 Oy 375 TleGlyGlyGlyGlyLe-----GlyPheThrile 384
 Db 1686 GTGGGACTGTAAGGAATTCGAAACAAAGCTGATTCAGATACCTCAAGAGCAACCTTTC 1745
 Oy 385 TyrGln-----GluIleIleThrGlyLe 392
 Db 1746 TCAACAAACGCTTCTTAATATACGATTCGTAATACATATTAGAAAGATTTCTCGCGAC 1805
 Oy 393 GluProAlaLeuThrProAlaGluIleGlyThrAspAlaLeuProIleLeuAlaLeuLys 412
 Db 1806 GGGATTGCTCTGTGCAGAACCAAGCTGCGGTGTCGGAAGGAATTTGCTGTCGA 1865
 Oy 413 AspIleGluAlaGluTyrLysSerGlyLeuAsnValGlyTyrSerIleTyrAsn 432
 Db 1866 ATGTGCATGCCGATTTATGTTACACAAATATACGGCGGCTTGGCAATGCTGTAGAT 1925
 Oy 433 ThrGlyGlyGlyGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyGlu 452
 Db 1926 GACGACCTGTTCCAAATATACGAAGTATACGAATCAATATGATTTACAGATGCTGA 1985
 Oy 453 ValIleGluValLeuThrSerSerIleAsnIleSerTrp---TyrGlyTyrGlyGlnHis 471
 Db 1986 GCTTATGATGTGACATGACATGCAACAAATGTTATTCATCATATTCACATTCAT 2045
 Oy 472 ProGlySer-----PheSerLeuAlaProAsnGlnLeuSerGlnGlyLe---AsnThrIle 489
 Db 2046 AATACTGGCGGCTTTTATGCTTATGCTGTGATCAAAAAAATCGGTTATTCGGTCAAT 2105
 Oy 489 eThrLeuLeuTyrArgArgThrGlyThr---GluGlnTrpGluProValArgHisAlaGlu 508
 Db 2106 ATCTCACTAATATGACGGCGGTGTATCCAGGACAGGATTAAGCAATCCAGGC-GGACC 2164
 Oy 508 nGlyGlyTyr-----ValAsnSerIleLeu 516
 Db 2165 TGTGTGCTCAATTAATTAAGAACAAAGCATTTTCCATTAATTTGGTGAAAAAATGATGGAC 2224
 Oy 516 eValAsnThrThrAspProAsnAsnValValValThrValAsp----- 530
 Db 2225 GGCATATGCCACTATCCACAAATATATAGCATTTATGTCGACGGATTTTCAACATCAT 2284
 Oy 531 -----AsnAsnGlyLeuSerIleValProAsnSerPh 543
 Db 2285 ATTTGCTGAAGAAATAGTTCGACACACGCTGTACAGTACCAATTTTACATATATAT 2344
 Oy 543 eValAlaAspLeuAsnSerTyrGlnHisSerThrIleThrValGlnPheAsnSerAspSe 563
 Db 2345 TTTA-----TACAAAGAAAGTACCGGACAAATTAATTCCTATCTAATTA 2389
 Oy 563 TrpAspGluIleArgThrProValAlaPheAlaLeu----- 575
 Db 2390 TCCAAACAAATGGGACACAAACCAATGAAATGTCGATTAATCCAGAAAAATATTT 2449
 Oy 576 -----SerThrGlyAlaIleThrAlaAs 582
 Db 2450 CAAAAATATATTATTTGGCCAAAAAGAAATTCCTACAGAAAAAGGTGGTCACTTTGA 2509
 Oy 582 pAspValIleSerLeuGlyTyrPValMetAlaGluValProGlyGlySerSerAsnTyrPr 602
 Db 2510 AAAATATAGATCAAGGAAACATTTTCAAAAAACACAAATTAAGAAATCAGATTAACCC 2569
 Oy 602 oValValTrpSerLysAspValLeuThrLeuSerGlnGlyAspTyrThr----- 618
 Db 2570 T-----GAAAGAGTGAAGAGTTGGCAGAGCAAAATTAACGACATTTGAACCC 2617
 Oy 619 -----LeuTrpTyrArgPheSerIle 625
 Db 2618 AACAAAGATATATGTTGATTGAAATTTACTTCTAAAGAACGCTTACGTCAGCGACACAAAT 2677

Oy 625 eAsnAsnGlnLysAspGluTrpLysIleGlySer-----ValSerValIleThrPr 643
 Db 2678 GTTCCGATTAATAAGAAACCTCCACAGCTTGTGTAAGCTGTGCGAAGTGAACAGTCC 2737
 Oy 643 oThrGlu-----TyrThrHisProIlePheGluValGlyHisAsnGlu 657
 Db 2738 GCCCGAGACTCTTTGGAATATAGCTTGAATAATTAAGTTTATGATATGTCGCCACA 2797
 Oy 657 nThrSerThrTyrThrLeuAspMetAlaHis---AsnArgValIleLeuProAspPheThrile 676
 Db 2798 ACCATCCACAAATGAAAAAGATTCGTTATCAAAATCAAGCTTGAATAATAGTTCCGCC 2857
 Oy 676 uLysAsnLeu-----GlyLeuProPheAsnGlyLeuValVal-----ValPheArg 692
 Db 2858 AACAGTGTGTATCCCAATTTGTCAGAGCAAGTGAATTAACGTACGAGAAAGTCTTAA 2917
 Oy 692 gGlnThrGlnSerSerSerGlySer-----LeuTrpAlaIle-----Glu 705
 Db 2918 TGAAGAACTGTTGGCAGACAGCAAAAAAGAAATTAATTAATGATTAACATGGAGCAATCCCTCA 2977
 Oy 705 nGluThrValHis-----IleLysGlnGlyGluThrPheValTyrLysProVal 721
 Db 2978 AGAAAAAATTAATATACAGACAGCACAGTGAAGTTCGACAGACAGTATTCGCTTACCGAT 3037
 Oy 721 I-----ValGlyGlyProIleProAspGlySerTyr 731
 Db 3038 AGATGCTGTCAAGTACAGCAGCAAGCTCTTTTGAAGGTGAAGTTCGTAAGGAAGAAACA 3097
 Oy 731 rArgAlaThrLeuHisAlaPheValAsnGlyGlnGlnLeuTyrLeuLys----- 748
 Db 3098 CACCGTTAACTCAAAACAA-----GCTCAAAAGTCCATATGTTCAAAAGAGTGA 3145
 Oy 749 -GlyLysArgAsnTyrThrValLysIleValAsnGlyThrAlaValGluAlaIleGluSe 768
 Db 3146 TGTATACCGACGATCTGTGCT---ATACAGCGGTACCGCT-----GCCATTAGTTC 3196
 Oy 768 rSerGlnGluIleArgValPhePro-----AsnProAla----- 779
 Db 3197 AGGAGACGCGCTTAATA---TATCCTTATGAGTAACATATACGGCAATTAACGATTA 3253
 Oy 780 -----ArgAspTyrValGluIleSerAlaTr 788
 Db 3254 ATTAATAATGCTAGTCTCGCGGTATATATCGCGGATTTACGTAATTAATGATACACA 3313
 Oy 788 oCysIleProGlnGluThrSerIleIleLeuPheAspLeuSerGlyLysIleValMetLys 808
 Db 3314 AGAATGAAGAACTATCAATCACTAATAAAGTGCATTTTGAACATTAAT---GTGATGCG 3370
 Oy 808 sAsnSerLeuSerAlaGlyHisGlyArgMetAspValSerArgLeuProAsnGlyAlaTyr 828
 Db 3371 TATATGCTGGTGAATAGGTGCTGGAATATCAATATGATGAACGTTGAAAAAATACGACAGT 3430
 Oy 828 rIleLeuLysValAspGlyTyrThrThrLysIleAsn 840
 Db 3431 C-----CAAGTATGATTTATGAAACAGATACTAAT 3461
 Db
 RESULT 20
 ID AAS75934
 XX AAS75934 standard; cDNA; 9466 BP.
 AC AAS75934;
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #11738.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 PN WO2001/5067-A2.


```

Db 6282 GACCTGGAAGANGGACCTGCAAAAGTCTCTACTTCCCTACCGCTGGCTGGTTATATC 6341
Qy 369 -----SerProthriserleuglyle 376
Db 6342 GTCTCCACCAAAATTCCTGACGACGACGTCCCTGGAGCCCATTTACCTGAAGATCAGT 6401
Qy 377 GlyGluGlyLeuGlyPheThrIleTyrGlnGluIleIleThrGlyLeuProAlaLys 396
Db 6402 GGGGAGGGA-----AGAGTCAAAAGAGCATCACCCGACACGATGGGCCCG 6449
Qy 397 ThrProAlaGluAlaGlyThrAspAla-----LeuProIleLeuAlaLeu 411
Db 6450 TCCGTGGCCACGTCTGGAGCATTTGTGACCTGAACCTGAATAATCCAGAAACAACAC 6509
Qy 412 LysAspIleGluAlaGluTyrLysSerGlySerGlyLeuAsnValGlyTyrSerIleTyr 431
Db 6510 AGGTATATGTCGGCCACGTCACACCCCTCTGGC----- 6545
Qy 432 AsnThrGlyGluGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGly 451
Db 6546 -----CGTGTGACTGAGGCA----- 6560
Qy 452 GluValIleGluValLysThrSerSerIleAsnIleSerTyrGlyTyrGlyGluHis 471
Db 6561 GACATTGTGCCATG-----GGGAGAAC 6584
Qy 472 ProGluSerPheSerLeuAlaProAsnGlnLeuSerGlnGlyLeuAsnThrIleThrLeu 491
Db 6585 TCACACTGGCTCGGTTTGTGTCGCCAGAGATG-----GGCGTGACACGGTCACGTC 6638
Qy 492 LeuTyrArg----- 494
Db 6639 AACTACCGTGGGAGCAGCAGTCACCGGACCCCTTCAGTTCACCGTGGGCGACCTGGT 6698
Qy 495 -----ArgThrGlyThrGlnGlnTrpGluPro--- 503
Db 6699 GAAGGAGGCGCCCAAGGTCGGGCGAGGAGGCCCTGGCTGGAGAGAGAGAGAGCGGGA 6758
Qy 504 -----ValArgHisAlaGlnGlyTyrValAsnSerIle 515
Db 6759 GTCCAGACGTAGTTACGATTTGACCCGGGAAGCGGCTGGA-----GGCCTC 6809
Qy 516 LysValAsnThrThrAspProAsnAsnValValIleThrValAsnAsnGlu---Gly 534
Db 6810 TCCATGCGTGTAGAGGCCAGTAAAGCCGAGATTACATTGATGACCAATAAAATGGG 6869
Qy 535 LysLeuSerIleValProAsnSerPheValAla---AspLeuAsnSerTyrGlnHisSer 553
Db 6870 TCGTGGGGTGA-----TCTTATATGCCCCCAAGAGCCCTGTACTACGAG----- 6914
Qy 554 ThrIleThrValGlnPheAsnSerAspSerProAspGluIleArgThrProValAlaPhe 573
Db 6915 ---GTGTCCATCAAGTTCAAT-----GATGAGCACATCCCGGAAGCCCTTAC 6959
Qy 574 AlaLeuSerThrGlyAlaThrAlaAspAspValIleSerLeuGlyTyrValMetAlaGlu 593
Db 6960 CTGTGTCGGGTATCTACACCTCCGACGAGCGCCGCCCTGACCTTATAGACCTTCAG 7019
Qy 594 ValProGlyLysSerSerAsnTyrProValIleTyrSerLysAspValLeuThrLeuSer 613
Db 7020 GAATCGGGAATTAAGTAAACCAAGCCAGCTCC----- 7052
Qy 614 GluGlyAspTyrThrLeuTyrTyrArgPheSerIleAsnAsnGlnLysAspGluTyrLys 633
Db 7053 -----TTTGCTATTAAGTTGATGAGCGCAAAA----- 7079
Qy 634 LysIleGlySerValSerValLysThrProThrGlyTyrThrHisProLeuPheGluVal 653
Db 7080 -----GGCAAGATTGATGCAAGGTCACAGC-----CCCTCTGGAGCGCTG 7121
Qy 654 GlyHisAsnGlnThrSerThrTyrThrLeuAspMetAlaHisAsnArgValLeuPro--- 672
Db 7122 GAGGAGTCCACGTGTCTGAGTGGAGCCAGATAGTATGCTGTGCTTACCTCCAT 7181

```

```

Qy 673 AspPheThrLeuLysAsnLeuGlyLeuProPheAsnGlyGluLeuValIle----- 689
Db 7182 GAGATGTCGTCCACACCATCATGATGCAAGTTCAATGGAGCCAGCTGGTGGAGCCCC 7241
Qy 690 -----ValPheArgGlnThrGlnSerSerGlySerLeuThrAlaAlaGln 705
Db 7242 TTCAAAGTCGGGTGGGAGGAGCCGTGACAAAGCGGGGAACCCCTGTGTCCCTCAT 7301
Qy 706 GluThrValHisIleLysGlnGlyLeuThr-----PheValTyrLysPro 720
Db 7302 GGCAGG---GGACTGAAAGGGGCGACACAGGTATTCAGTCCGAATTTCTTTATACACC 7358
Qy 721 ValValGluGlyProIleProAspGlySerTyrArgAlaThrLeuHisAlaPheValAsn 740
Db 7359 ACCGAGCAGGTGCA-----GGGACATTTATCCGTACACATCGAA 7397
Qy 741 Gly-GlnGlnGlnLeu-----TyrLeuLysGlyLysArgAsnTyrTh 754
Db 7398 GGCCCATCCAAAGTTAAATGATGATGCCAGAAACACCTGAAGGGGTACAAAGTCATTAC 7457
Qy 754 rValLysIleValAsnGlyThrAlaValGluAlaIleGluSerSerGluGluIleArgVa 774
Db 7458 ACCCCCATGGCTCGTGGTAACTACTGATCACTGTCAAATACGGTGG----- 7506
Qy 774 ILeProAsnProAlaArgAspTyrValGlu----- 784
Db 7507 ---CCCAACCAACATCGTGGCAAGTCCCTTCAAGGCCAAGTGACAGGCCAGCT 7562
Qy 785 ILeSerAlaProCysIleProGlnGlnThrSerIleIleLeuPheAspLeuSerGlyLys 804
Db 7563 AGTTAGCCCTGG-CTCAGGCCAAAGCAGACCTCATTCATCTGTGTGAGTCACTGACCA 7621
Qy 804 sIleValMetLysAsnSer 810
Db 7622 GGTCTCTCAAGAGACT 7640

```

Search completed: June 1, 2003, 22:41:03
 Job time: 633 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OW protein - nucleic search, using frame_plus.p2n model

Run on: June 2, 2003, 00:16:06 ; Search time 301 Seconds

(without alignments)
3776.318 Million cell updates/sec

Title: US-10-030-330-1

Sequence: 1 MKKSFILAVIMFGIAMCGH.....PNCAYILKVDGYTKRINIVH 843

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 845702 segs, 674182571 residues

Total number of hits satisfying chosen parameters: 1691404

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame.p2n.model -DEV=ylh
-Q=/cgn2_1/USPTO/spool/US10030330/runat.23052003.181646.6824/app_query.fasta.1.1031
-DB=Published_Applications_NA -QMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-FRAN=human40.cdi -LIST=45 -DOCLIGN=200 -THR.SCORE=pct -THR.MAX=100
-THR.MIN=0 -ALIGN=20 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEARST=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10030330.ecgn.1.1.170.0runat.23052003.181646.6824
-NCPU=6 -ICPU=3 -NO.XLPRY -NO.MAP -LARGEQUERY -NEG.SCORES=0 -WAIT
-DSBLOCK=100 -LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10
-XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications_NA:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PTCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	403.5	9.1	1419	US-10-002-784A-23	Sequence 23, Appl
2	154	3.5	3878	US-09-741-669-111	Sequence 111, App
3	148.5	3.3	4059	US-09-784-554B-1	Sequence 1, Appl
4	148.5	3.3	9047	US-09-070-927A-12	Sequence 12, Appl

5	145	3.3	25165	US-10-114-170-39	Sequence 39, Appl
6	140	3.2	3106	US-08-837-459-21	Sequence 21, Appl
7	139.5	3.1	3695	US-10-114-170-18	Sequence 18, Appl
8	138.5	3.1	4605	US-10-080-505-10	Sequence 10, Appl
9	138.5	3.1	4822	US-10-080-505-14	Sequence 14, Appl
10	138.5	3.1	10455	US-10-160-758-4	Sequence 4, Appl
11	138	3.1	4765	US-10-090-624-5	Sequence 5, Appl
12	138	3.1	4828	US-10-080-505-16	Sequence 16, Appl
13	138	3.1	9299	US-09-848-616-169	Sequence 169, App
14	138	3.1	9299	US-09-848-616-168	Sequence 168, App
15	137.5	3.1	4268	US-10-114-170-93	Sequence 93, App
16	136.5	3.1	11065	US-10-160-758-1	Sequence 261, App
17	136.5	3.1	11073	US-10-160-758-2	Sequence 20, Appl
18	133.5	3.0	7712	US-09-070-927A-489	Sequence 2, Appl
19	133	3.0	4557	US-09-874-300-2632	Sequence 2632, App
20	133	3.0	43360	US-10-114-170-206	Sequence 206, App
21	133	3.0	45345	US-10-114-170-261	Sequence 261, App
22	132.5	3.0	3131	US-08-837-459-20	Sequence 20, Appl
23	131	3.0	8979	US-09-738-626-2739	Sequence 2739, App
24	131	2.9	3309400	US-09-738-626-1	Sequence 1, Appl
25	127.5	2.9	3309400	US-09-738-626-1	Sequence 1, Appl
26	126.5	2.9	5629	US-10-114-170-243	Sequence 243, App
27	125	2.8	2637	US-09-912-020-204	Sequence 204, App
28	125	2.8	5010	US-09-917-800A-483	Sequence 483, App
29	125	2.8	5681	US-09-848-616-164	Sequence 164, App
30	124	2.8	4056	US-09-784-554B-3	Sequence 3, Appl
31	123.5	2.8	7198	US-10-114-170-147	Sequence 147, Appl
32	122.5	2.8	2604	US-09-839-894-5	Sequence 5, Appl
33	122.5	2.8	7239	US-09-839-894-27	Sequence 27, Appl
34	122	2.7	6702	US-09-769-787-209	Sequence 209, App
35	121	2.7	7035	US-09-815-242-8615	Sequence 8615, App
36	120.5	2.7	5245	US-10-080-505-12	Sequence 12, Appl
37	120	2.7	1791	US-10-213-890-50	Sequence 50, Appl
38	120	2.7	2870	US-10-121-032-59	Sequence 59, Appl
39	120	2.7	2870	US-10-093-037-59	Sequence 59, Appl
40	120	2.7	3169	US-10-131-591A-2	Sequence 2, Appl
41	120	2.7	3597	US-09-975-719-404	Sequence 404, App
42	120	2.7	17388	US-09-815-242-8512	Sequence 8512, App
43	120	2.7	42235	US-09-975-719-1	Sequence 1, Appl
44	119.5	2.7	4794	US-10-092-880-7	Sequence 7, Appl
45	119	2.7	6091	US-10-114-170-200	Sequence 200, App

ALIGNMENTS

RESULT 1
US-10-002-784A-23
; Sequence 23, Application US/10002784A
; Publication No. US200303644A1
GENERAL INFORMATION:
/33
APPLICANT: Ulrich, Robert G.
FILE REFERENCE: Bacterial Superantigen Vaccines
CURRENT APPLICATION NUMBER: US/10/002,784A
CURRENT FILING DATE: 2001-11-26
PRIOR APPLICATION NUMBER: 08/882,431; 09/144,776
PRIOR FILING DATE: 97-06-25; 98-09-01
NUMBER OF SEQ ID NOS: 40
SOFTWARE: Apple Macintosh Microsoft Word 6.0
SEQ ID NO 23
LENGTH: 1419
TYPE: DNA
FEATURE:
ORGANISM: Artificial sequence
OTHER INFORMATION: mutant Spea/mutant Speb fusion
US-10-002-784A-23

Alignment Scores:
Pred. No.: 7.49e-33 Length: 1419
Score: 403.50 Matches: 105
Percent Similarity: 50.17% Conservative: 47
Best Local Similarity: 34.65% Mismatches: 114


```

1372 GGAACCCCAATATTAGGATGGCGGTAGTTCGATCGCAAAATATTACTTT----- 1425
OY 332 AsnTrpGlyTrpGlyValSerAsnGlyPheTrpLysLeuThrLeuLeuSerProThr 371
Db 1426 -----TTGGCGCATAAAGCAAGCGGTATCGCACACTGATCGATGATGCGACA 1476
OY 372 Ser-----LeuGlyIleGlyGlyGlyLeu----- 380
Db 1477 TCATCGTTCGATCTGGGTATCATGTCTGGTAAATTTGGTAGCGTATCGTAAATGTC 1536
OY 381 -----GlyPheThrIleTrpGlnGlu-----IleIleThrGlyIleGluProAla 395
Db 1537 AGTATGGTGGCACCCCTTAATTCACAGCGGCTATGATTTATCGAGGAATCCCTCCGGT 1596
OY 396 LysThrProAlaGluAlaGlyThrAspAlaLeuProIleAlaLeuLysAspIleGlu 415
Db 1597 AAGGGAGATGTAATATTCAACGCGACGCTCTGGAATTTAAACATCATCTACTAAC 1656
OY 416 AlaGluTrpLysSerGlySerGlyLeuAsnValGlyTrpSerIleTrpAsnThrGlyGlu 435
Db 1657 GCACATATG-----CTACAGCTCGGT-----GTAATAGGCGCGGGTGA 1695
OY 436 GluGlnSerAsnLeuAspLeuGlyTrpArgLeuAsnLysAlaAspGlyValIleGlu 455
Db 1696 -----GluProVal-----CTGATATTACACCGGAGGTACTTAA 1725
OY 456 ValLysThrSerSerIle-----AsnIleSerTrpTrpGly 467
Db 1726 GCGCGTGATACACAGATAGCTCTCAATGACAAATAGAGGCGGTAGGCGGTGATGG 1785
OY 468 TyrGlyGlnHisProGlnSerPheSerLeuAlaProAsnGlnLeuSerGlnGlyIleAsn 487
Db 1786 CAGAACTCTCTTTGTAACATCATATGATGATAGG-----ACATCTGCTACGGGT 1839
OY 488 ThrIleThrLeuLeuTrpArgTrpThr-----GlyThrGlnGlnTrp----- 501
Db 1840 ACGTTAACCTGACGATACAGGATACGCTGATGATGCGAAGGTGAGGATTTACTTAGGT 1899
OY 502 -----GluProVal-----ArgHisAla 507
Db 1900 GTTTTGTAGCTGCTGTAGGAACGCTAAACATTTGGTCTGCTCAGCGTAGCGCGCGCA 1959
OY 508 GlnGlyGlyTrpVal-----AsnSerIleLysVal----- 517
Db 1960 GATCGCGATTTATCACCAATGCGACGAAAGTGAGTTGGTCTTGGCGAAGCGCTTTT 2019
OY 518 -----AsnThrThrAspProAsn-----AsnValValValThrVal 529
Db 2020 GTCTTTAATCATACCAATACAGATGATCGCGCTACGAGTCGATATGCTGATACAGGT 2079
OY 530 AspAsnAsnGlnGlyLysLeuSerIleValProAsnSerPheValAlaLysLeuAsnSer 549
Db 2080 GACGATAAAGACGGAAAGTGATC----- 2103
OY 550 TyrGlnHisSerThrIleThrValGlnPheAsnSerAspSerProAspGluIleArgThr 569
Db 2104 -----CATGATGACAGGCCATACGCTGTTCAATGACGGAATATTAAAGCGTAAACG 2157
OY 570 ProValAlaPheAlaLeuSerThrGlyAla-----ThrAlaAspValIleSerLeu 587
Db 2158 CTGTCATGACGCGCTCTCGCATTCGCTCTCATGAGGCGAGATGGGTTAACGGCGCATG 2217
OY 588 GlyTrpValMetAlaGluVal-----ProGlyGlySerSerAsnTrpProVal 603
Db 2218 GGG-----TCGAGTGAAGTACCAATTCGCAAAACCCCGGT----- 2250
OY 604 ValTrpSerLysAspValLeu-----ThrLeuSerGlnGlyAspTrpThrLeuTrpTrp 621
Db 2251 -----ACGCTGACATTCCTGCATCAACGAACAGTGCAGAGATTTACACGCTGACCAAT 2304
OY 622 ArgPheSer-----IleAsnAsnGlnLysAspGluTrpLysIleGlySer 637

```

```

Db 2305 GCGCTCAAGGCGATGCGCTTATGCGAGTGCAGCTCTCATCTCCGACAAAGATGTTGGC 2364
OY 638 ValSerValLysThrProThrGluTrpThrHisProLeuPheGluValAlaHisAsnGln 657
Db 2365 TTATCCACATGACAAAGGCGCTGAAATTCGCC-----GGTGTCCCAACTGAA 2412
OY 658 ThrSerThrTrpThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrIleLys 677
Db 2413 GACAGACCTTCATCTTGGACCGCGCAACACCCCTGCGCTTACTACCGGATGTTGCA 2472
OY 678 AsnLeuGlyLeuProPheAsnGlnGlyLeuValValPheArgGlnThrGlnSerSer 697
Db 2473 TCT-----GACAGTGAATAATACACATCGGTAAATTTGGAGAGCAATCC 2517
OY 698 SerGlySerLeuTrpAlaAlaGlnGluThrValHisIleLysGlnGlyGluThrPheVal 717
Db 2518 ATTGGTGGACTG-----GCCATGAAATGAGATGACCAATCAT 2553
OY 718 Tyr-----LysProValValGlnGlyProIleProAspGlySerTrp 731
Db 2554 TTGCGATACGGAATTCCTCTCGACGCTCGCGGAGGATATATC----- 2598
OY 732 ArgAlaThrLeuHisAlaPheValAsnGlnGlnGlnLeuTrpLeuLysGlyLysArg 751
Db 2599 -----ACGCTGATACGCTGCTGTTGCGCGCGGCGGTGATACACCTGAAAGGC---CGT 2649
OY 752 AsnTrpThrValLysIleValAsnGlyThrAlaValAlaIleGluSerSerGlyGlu 771
Db 2650 AACTATCAG-----GTAACGCGGACGCGGCGACGCTTATCAGC----- 2688
OY 772 IleArgValPheProAsnProAlaArgAsp 781
Db 2689 -----GTGCTTAACCGTGGAAATGAT 2709

RESULT 3
US-09-784-554B-1
; Sequence 1, Application US/09784554B
; Publication No. US20030032162A1
; GENERAL INFORMATION:
; APPLICANT: Schmitt, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schuelin, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGUCANASES
; FILE REFERENCE: 10017, 200-US
; CURRENT APPLICATION NUMBER: US/09/784, 554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 4059
; TYPE: DNA
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-1

Alignment Scores:
Pred. No.: 3 56e-05 Length: 4059
Score: 148.50 Matches: 176
Percent Similarity: 32.10% Conservative: 127
Best Local Similarity: 18.64% Mismatches: 334
Query Match: 3, 35% Indels: 307
DB: 9 Gaps: 47

US-10-030-330-1 (1-843) x US-09-784-554B-1 (1-4059)
OY 112 AspMetAspSerMetProAspAsnLeuArgMetTrpLeuGlnIleTrpAspGlnIle 131
Db 844 GATGGAGCTCTGTAAAGCAATATATACCTGTTCTGACATATATACCTGATCAATG 903
OY 132 GlyLeu----- 133
Db 904 GCGCTTACCTGCAAGTCGACAGCAAGAGATTCGATGATTCGACGCTACAGTGTAT 963
OY 134 -----IleLeuSerGlyLysAlaGlnLeuAsnGlnIleLeuArgThrGlnGlyVal 151

```

```

Db      964 CCCGAGCGATGGGGAGGAGCATACGAATTACGAATGAGTAGGCAATGACAAACGAAG 1023
Qy      152 ProAGLValAlaLeuMetAspAsnGlyHisPheAlaAsnAspProMetArgTrp 171
Db      1024 AAGCCAGAAATGCAGCA-----CCTGGCACTTTGTGG 1056
Qy      172 AsnGlnGlyTrp-----ProTrpAsnAsnLys---GluProLeu 183
Db      1057 GACCCGACCTATAGGAAGATAGTTGATCGCTCAATGAGAACAGGATTTTGGCCANA 1116
Qy      184 LeuPro----- 185
Db      1117 CTACCTCGATTGAGACAGCTGGTGATAAATATTATCCGGAGCAACAGCTGGCAATGACC 1176
Qy      186 -----AsnGlyAsnHisAlaLeuTrpGlyCysValAlaThrAlaAlaAla---- 200
Db      1177 GAGTATAGCTATGCGGGCAAAATGATATTTCCGGCGGATTCGCGATGACCATGGCTG 1236
Qy      201 -----GlnIleMetArgTrpHisSerTrpProLeuGln---GlyGlu 213
Db      1237 GGTATCTGGGCCAAATGATGTTTATATGCAAACTACTGCAAGCTTAAGAGATGGTGC 1296
Qy      214 GlySerPheAspTrpHisAlaGlySerLeuValGlyAsnTrpSerGly-----Thr 230
Db      1297 AACAACTACGTTATAGTCCGCTTACAACTTTATCCGAATATGACGAAATAAATCTTACT 1356
Qy      231 PheGly-----GluMetTrpAspTrpIleAsnMetProGlyAsnPro 244
Db      1357 TTCGGTGATACCAAGTGTAGTGCAGCAACATGCGATATGTCAATATGCTCGGTCCATGCT 1416
Qy      245 AspLeuAspAsnLeuThrGlnSerGlnValAlaAspAlaTrpAlaThrLeuMetArgAspVal 264
Db      1417 TCTGTACAGCAATGATCCGACAAAGAACTGCAT---CTCGTGTGCAATGAAAGACATG 1473
Qy      265 SerAlaSerValSerMetSerPheTrpGluAsnGlySerGlyThrTrp---SerValTrp 283
Db      1474 GACACGCGCTTGCAGCGCCCAATTTGATCTTCCGGCGGAGAGACTTACCTTCGGGTAA 1533
Qy      284 ValValGlyAlaLeuArgAsnAsnPheArgTrpGlyArgSerLeuGlnLeuHisVal--- 302
Db      1534 GATATGGGCTTGCATTAATAAACAGCTCGCAATTTMAAGAACAGCCGCAATACGCAAT 1593
Qy      303 -----ArgAlaLeuTrpThrSerGlnGluTrpHisAspMetIleArgGlyGluLeu 319
Db      1594 TCAGGCAACCCGTTTACTTATACCTACGTCAGC-----CCTTGG 1629
Qy      320 AlaSerGlyArgProValTrpTrpAlaGlyAsnAsnGlnSerIle-----GlyHis 336
Db      1630 ACGGCAATATCACATTTGCTGCTACTGCTACGTGCAATGACAGCTCTCCAGTGGAGGTCCTGAA 1689
Qy      337 AlaPheValCysAspGlyTrpAlaSerAspGlyThrPheHisPheAsnTrp----- 353
Db      1690 AGCTTTAAGCTGAAGAGCTGAGCTGTGATGGGAAAGTCCATTATTCCTGGGATGCTTCC 1749
Qy      354 -----GlyTrpGly----- 356
Db      1750 AGCGAGTGTAGGATACAGCGCTACAGCGGCGCAACAGATGAAGAGGCCCTTTCATCTGCT 1809
Qy      357 -----GlyValSerAsnGly----- 361
Db      1810 GTAGCATCCAACTTACCGAAACGCTTATACAGGATACCTAACGTGCAAAACGGTACTCTCA 1869
Qy      362 ---PheTrpLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyLysGluGlyIle 380
Db      1870 TACTATTACAAAGTAAC-----GCCAAACCAATTAAGGATCGAGCAATCCCAATATT 1923
Qy      381 GlyPheThrIleTrpGlnGluIleIleThrGlyIleGluProAlaLysThrProAlaGlu 400
Db      1924 TTGAAGCGTTCCGAGATGCGCTGTAAGCGT-----CCGCTGCTATGACGCGGA 1977
Qy      401 AlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAsp-----IleGluAlaGlyTrp 418

```

```

Db      1978 GAAGCAGC-----CTGAAGGCAACCATTTGTGGAATCCAGCGG 2016
Qy      419 LysSerGluSerGlyLeuAsnValGlyTrpSerIleTrpAsnThrGlyGluGlnSer 438
Db      2017 ACCGGCTACTCCGGTCCGTGCTTATGACGAATTTCCAAATCCAGGGATTCCTGACG 2076
Qy      439 -----AsnLeuAspLeuGlyTrpArg----- 445
Db      2077 ATGACGATTACAGCTCCACGAGCGCTGTGCAATCTTACAAATGCGGTACCGCTCTCTCT 2136
Qy      446 -----LeuAsp---LysAlaAspGlyValIleGlu 455
Db      2137 CATGATGACAAACGACCAATTTCTCATTAACGCGAAAGCGTTGGCGCATCGCTT 2196
Qy      456 ValLysThrSerSerIleAsnIleSerTrpGlyGlyGlyHisProGluSerPhe 475
Db      2197 AAGMAAACGCGTATTT-----AAAGAACT 2223
Qy      476 SerLeuAlaProAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuLeuTrpArg 495
Db      2224 TCCGAGAGCGAAGTGTCTGTGAATGACGCGCGGAATACAGTACGATGTTT-----GAA 2274
Qy      496 ThrGlyThrGlnGlnTrpGluProValArgHisAlaGlnGlyLysTrpValAsnSerIle 515
Db      2275 ACAGGCTGG---GGCTGTGACATATC-----GACTACGTCAGACTGGAG 2316
Qy      516 LysValAsnThrThrAspProAsnAsnValValThrValAspAsnAsnGlnGlyLys 535
Db      2317 CCTGCCGCTGACCGCCGACCTCATGGGGAACCAAAAGCCTTACCAAT----- 2364
Qy      536 LeuSerIleValProAsnSerPheValAlaAsp-----LeuAsn 548
Db      2365 -----CCGAATGCGAGCGTGAAGCAAGCAAAAGCATGTGATCTGTGAT 2412
Qy      549 SerTrpGlnHisSerThrIleThrValGlnPheAsnSerAspSerProAspGluIleArg 568
Db      2413 CAATACGGGCAAGATATGCTCTGTGTCAGAGGAATTAACGAATTAATGATGGCTTCA 2472
Qy      569 ThrProValAlaPheAlaLeuSerThrGlyAlaThrAlaAspAspValIleSerLeuGly 588
Db      2473 GCCAATGTA-----GTAATAAGCGCGGATTCGACGCGCTTACCTGATCGACATATCG 2526
Qy      589 TrpValMetAlaGlu-----ValProGlyGlySerSerAsnTrpProValValTrpSer 606
Db      2527 CCAGACAGACGCGAACACGCTTGTATGTCACAGAGCGAAGAGCGATTCATGGGAT 2586
Qy      607 LysAsp-----ValLeuThrLeuSerGluGlyAspTrpThrLeuTrpTrp----- 621
Db      2587 AAGCAAGGGGGGATGCTTACCTTGC-----TGGCACTGGAACGCA 2628
Qy      622 ---ArgPheSerIleAsnAsnGln---LysAspGluTrpLysLysIleGlySerValSer 639
Db      2629 CGGAAGAGCTGATGATACGACGAGGAAGAAAGATGAGAGGCTTATCCGATTC 2688
Qy      640 ValLysThrProThrGluTrpThrHisProLeuPheGluValGlyHisAsnGlnThrSer 659
Db      2689 ACCACATTCGATATGAAATATCG-----ATGAATCATCCAGATCCGCA 2733
Qy      660 ThrTrpThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLysAsnLeu 679
Db      2734 GATTTAATTAATTAATTCATTTGCGACATCCATGATGATTCAGG---CAATTGAAGAAGTTG 2790
Qy      680 GlyLeuProPheAsnGlyLeuLeuValValPheArgGlnThrGlnSerSerSerGly 699
Db      2791 CAG-----GATGCGAAGTCTCTGTCGCTTCGCTTCCTTGCACGAACGCGAAGCA 2841
Qy      700 Ser-----LeuThrAlaAlaGlnGluThrValHisIleLysGln----- 712
Db      2842 AATGCTTCTGTGGGGCGCCAAAGTCCCTGAGCCTGTAAAGAACTGATATTTTAATG 2901
Qy      712 ----- 712
Db      2902 CACGACCGTTTGACGAATGTGCACAAATGAACTGATTTGGGTATGGAATCTCTGT 2961

```

QY 713 -----Glycylthrethreval----- 717
 Db 2962 GCTCCGATGTGATCCGGAGACAGATGTGATATTGAGCTTGACTTATCCG 3021
 QY 718 -----TylyProvalVal----- 722
 Db 3022 CAGCAGGTGATTACAGCCCGCAATTTCAAAATACGAAGACCTTGTCATTGGCAG 3081
 QY 723 -----GlyglyProleProasp-----Gly 729
 Db 3082 GACAAAAAGTAGTTCATGACGAGAAATGACGACCGATCCGACCTGATTGATGANG 3141
 QY 730 SerTyraAlaThLeuHisAlaPheValAsnGlyGlnGlnLeuTyrlleuysGly 749
 Db 3142 GCGTATCAGCTCATTTGAGCTGAGTTCCTACATGATGAGATTTGTCGAGACGCGC 3201
 QY 750 LysArgAsn-----TyThrVallyslleValAsnGlyThraValAlaGluAlaile 766
 Db 3202 AAACAAAACAGCCTTGAGCATCTGAAAAAGTGTATATCATCCGACGCTCATTCGCTG 3261
 QY 767 GluSerSerGluGluIleArgValPheProAsnProAlaArgAspTyraValGluIleSer 786
 Db 3262 GATGAG-----CTCCAAACGAATTAAAGATGACGATTAAGTACTGAG 3303
 QY 787 AlaProCysIleProGlnGluThrSerIleleuPheAspLeuSerGlyslleVal 806
 Db 3304 CAGCGCTCCGATCCGGAGCTTCACGCTTCGAGCTGCGGCTGAAACGCGCAAGATCG 3363
 QY 807 MetLysAsnSerLeuSerAlaGlyHisGlyArgMetAspValSerArg-----LeuPro 824
 Db 3364 CTAAGCTGACAGCATCGCGCAATGCGAAAGCTATGAAAGTGAAGCTTCGACGACTGAA 3423
 QY 825 AsnGlyAlaTyrr 828
 Db 3424 AACGCGCGCTC 3435

RESULT 4

US-09-070-927A-12

Sequence 12, Application US/09070927A

Patent No. US20020120116A1

GENERAL INFORMATION:

APPLICANT: Charles A. Kunsch

Patrick J. Dillon

Steven Barash

TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 982

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44M storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/070, 927A

FILING DATE: 04-May-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/046,655

FILING DATE: 1997-05-16

APPLICATION NUMBER: 60/044,031

FILING DATE: 1997-05-06

APPLICATION NUMBER: 60/066,009

FILING DATE: 1997-11-14

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB369
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 9047 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 12:
 US-09-070-927A-12
 Alignment Scores:
 Pred. No.: 0.000121 Length: 9047
 Score: 148.50 Matches: 182
 Percent Similarity: 34.59 Conservative: 120
 Best Local Similarity: 20.85 Mismatches: 320
 Query Match: 3.358 Indels: 252
 DB: 10 Gaps: 44

US-10-030-330-1 (1-843) x US-09-070-927A-12 (1-9047)

QY 143 GluGluIleLeuArgThrGluGlyValProAlaGluValHisAlaLeuMetAspAsnGly 162
 Db 1074 GAAAAAGTAAATCAGCGAGATTTCACGACGCGAC---CATTTCTGCTG----- 1121
 QY 163 HisPheAlaAsnAspPrometaArgTrpAsnGlnGlyTyrrProTrpAsnAsnLysGluPro 182
 Db 1122 -----AMTGGCGCAAGTAACTCGAATTAAC---CAATG 1151
 QY 183 LeuLeuProAsnGlyAsnHisAlaTyrrThrGlyCysValAlaThrAlaAlaGlnIle 202
 Db 1152 TTGCATCCAAAAGGAAAT-----GGGACACGCGCCAAAGAT 1190
 QY 203 ---MetArgTyrrHisSerTrpProLeuGlnGlyGluGlySerPheAspTyrrHisAlaGly 221
 Db 1191 GTCATGATTTTATATATAGATGACAAAGGAAAGCAATTTTGAACACACGCGCT 1250
 QY 222 SerLeuVal-----GlyAsnTrpSer---GlyThrPhe----- 231
 Db 1251 CGGATATTAACGGTGGTGGACCTTACACACGAGGACTTCAACGTCGATCTCAGG 1310
 QY 232 -----GlyGluMetTyrrAspTrpIleAsnMet-----ProGlyAsn 243
 Db 1311 GCCGTGCAGTGTAGTGAACCAAGAAATTTGGGATTTTCAAAATTTAAGTGAACGAAATACA 1370
 QY 244 ProAspLeuAspAsnLeuThrGlnSerGlnValAlaPheAlaThrLeuMetArgAsp 263
 Db 1371 CCAGAGCTGATTAATCTC-----GAAGGCTATTAAGAAACGAGAGATGCT 1415
 QY 264 ValSerAlaSerValSerMetSerPheTyrrGluAsnGlySerGlyTyrrSerValTyrr 283
 Db 1416 CAAGAGCGGGGATATTAGTATAGATATGAAACAAATGCAACGTTCAATAGTGC--- 1472
 QY 284 ValValGlyAlaLeuArgAsnAsnPheAlaGlyTyrrLysArgSerLeuGlnLeuHisValArg 303
 Db 1473 -----ACGATTAAGAAATTAATCTGTTCAAGAT----- 1499
 QY 304 AlaLeuTyrrThrSerGlnGluTrpHisAspMetIleArgGlyGluLeuAlaSerGlyArg 323
 Db 1500 -----GTTCAACAGCAATATTTTAAATCTAGGCGGAATCTGCACCAACGA 1550
 QY 324 ProValTyrrTyraAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCysAspGlyTyrr 343
 Db 1551 -----TTAAAAGCGGTGAGTGAATATTTTATAGAACTGG 1589
 QY 344 AlaSerAspGlyThrPheHisPheAsnTrpGlyTyrrGlyGlyAla----- 358
 Db 1590 TTGATGAGAAAT-----GGCAAGTACTAAGCGACGACTAAT 1625
 QY 359 -----SerAsnGlyPheTyrrLysLeuThrLeuLeuSerProThrSerLeuGly 374

```

Db      1626 GATCATCGAACACAGCTGATTTAATGATATTTTGAATGAAAAATATGATTCACACGA 1685
QY      375 11eglygluglylle-----Glypethrile 384
Db      1686 GTGGACCTTGAGAGAAATTCGAACAAAGCTGATTCAGATACCTCAAGAGCCAACTCTTC 1745
QY      385 TyrGln-----Gluilelthrlyle 392
Db      1746 TACAAACGTTTCTAATATACGATTCGTAATACCTATTAGAAAGATATCTGCGCAG 1805
QY      393 GluProAlaLysThrProAlaGluAlaGlyThrAspAlaLeuProIleuAlaLeuLys 412
Db      1806 GGGATGCTCTTCACAGAACCAAAAGTGGCTGCTCGACAGAAATGCTCTCCGA 1865
QY      413 AspIleGluAlaGlyThrLysSerGlyLysLeuAsnValGlyTyrSerIleTyrAsn 432
Db      1866 ATTCGAAATGCCATATGTGTACACAAATACGGGGCGTTGGCAATGCTGTAGAT 1925
QY      433 ThrGlyGluGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyGlu 452
Db      1926 GACGACCTGTCATATATACAGACTATACGAAATCAAAATATGATTTAAGATGCTGA 1985
QY      453 ValIleGluValLysThrSerSerIleAsnIleSerTrp---TyrGlyTyrGlyGlnHis 471
Db      1986 CCTTATGATGTGACATGCATCAAAACATGTATATCATATATCACTATTCACATCAT 2045
QY      472 ProGluSer-----PheSerLeuAlaProAsnGlnLeuSerGlnGlyLe---AsnThrIle 489
Db      2046 AATACGCGGCTTTTATTCCTTATGCTGTATCAAAAATTCGGTTATTCGCTCATAT 2105
QY      489 eThrLeuLeuTyrArgThrGlyThr---GluGlnTrpGluProValArgHisAlaGlu 508
Db      2106 ATCCACCTATATGACGGCGGTGTACCGAGAACAGTAAGACATCCAGGC-GGAGC 2164
QY      508 nGlyGlyTyr-----ValAsnSerIleLys 516
Db      2165 TGGTGCTACATATTATAAGAACAAAGCATTTTCATATTGCGTGAAAAATGATGAGC 2224
QY      516 sValAsnThrThrAspProAsnAsnValValAlaThrValAsp-----530
Db      2225 GGCAATGCCACATATCCACATATATACGATTTATGTCGCGCATATTCACATCATAT 2284
QY      531 -----AsnAsnGlnGlyLysLeuSerIleValProAsnSerPh 543
Db      2285 ATTGGTGGAAGAAATAGTTCGACACAGTGTACAGTACCAATTTTACATATATAT 2344
QY      543 eValAlaAspLeuAsnSerTyrGlnHisSerThrIleThrValGlnPheAsnSerAsp 563
Db      2345 TTTA-----TACAAAGAAAGTACCGGACATTAATTAATCTTATCTAATTA 2389
QY      563 rProAspGluIleArgThrProValAlaPheAlaLeu-----575
Db      2390 TCCAACAAATGGACACACCAATTGAAAGCTTAATGCTGATTCAGAAAAATATTT 2449
QY      576 -----SerThrGlyAlaThrAlaAs 582
Db      2450 CAAAAATATATGTTATTTGCCAAAAAGAAATGCTACAGAAAAAGGTGCAACTGTGA 2509
QY      582 PAAPValIleSerLeuGlyTyrPValMetAlaGluValProGlyLysSerSerAsnTyrPr 602
Db      2510 AAAATTAAGTATCAAGGAAACATTTTGAACAAACCAATTAAGAAATCACAGATTAACC 2569
QY      602 oValIValTyrSerLysAspValLeuThrLeuSerGlnGlyAspTyrThr-----618
Db      2570 T-----GAAAAAGTAAAGAGTGGCAGACGACAGAAATTCAGACATTTGAACC 2617
QY      619 -----LeuTyrPyrArgPheSerIle 625
Db      2618 AACAAAGATATATGTTGTAATTTACTTCAAAAGACGCTTACGTCAGGACACACAAAT 2677
QY      635 eAsnAsnGlnLysAspGluTyrLysLysIleGlySer-----ValSerValLysThrPr 643
Db      2678 GTTCGATTAATAAGAAACTCTCCAGCAGATTGTAAGAGCTGTGCGAAGTGAACAGTCC 2737

```

```

QY      643 oThrGlu-----TyrThrHisProLeuPheGluValGlyHisAsnGlu 657
Db      2738 GGCCGAAGACTCTTTGGAAATAGCTGAAAAATTAAGATTTAGATATTTGGCCACACA 2797
QY      657 nThrSerThrTyrThrLeuAspMetAlaHis---AsnArgValLeuProAspPheThrLe 676
Db      2798 AGCATCCACAAATTTGAAAAAGATATTCCTTAATTTCCAGACCAAGTTGAATTAAGTAT 2857
QY      676 uLysAsnLeu-----GlyLeuProPheAsnGlyGluLeuValVal-----ValPheArg 692
Db      2858 AACAGGTGTGTATCTCAATTTTCCAGACCAAGTTGAATTAAGTATTAAGCAAGAGTCTTAA 2917
QY      692 gGlnThrGlnSerSerSerGlySer-----LeuThrPalaIle-----Glu 705
Db      2918 TGAAGAAAGTGTGGCAGACGCAAAAAAGATTTAAGTTACAAATGGGACCAATCCCTCA 2977
QY      705 nGluThrValHis-----IleLysGlnGlyGluThrPheValTyrLysProVal 721
Db      2978 AGAAAAATTAATATACAGACGACACAGTTGAAGTTGACGACAAAGTTATGCGCTTACCAT 3037
QY      721 I-----ValGluGlyProIleProAspGlySerTyr 731
Db      3038 AGATGCTGTCAAAAGTACAGCAAGGCTCTTTTGGAGGTGAACCTTGTCGAGAGAAAGA 3097
QY      731 rArgAlaThrLeuHisAlaPheValAsnGlyGlnGlnLeuTyrLysLys-----748
Db      3098 CACCGTTAACTAAAAACA-----CCTCAACTGCGCTATGTTCAAAAGAGTGA 3145
QY      749 -GlyLysArgAsnTyrThrValLysIleValAsnGlyThrAlaValGluAlaIleGlu 768
Db      3146 TGGTAAACCGACATCTGCT---ATACGACGCGTACCGCT---GCCATTAGTTC 3196
QY      768 rSerGlnGluIleArgValPhePro-----AsnProAla-----779
Db      3197 AGGAGAGCGCTTAAAT---TATCTTATGAGTAAGTAACATACGCGCAATTAACCATTA 3253
QY      780 -----ArgAspTyrValGlnIleSerAlaPr 788
Db      3254 ATTAATAATGCTAGTCTCGCGGGTATATACGCCGATTTACGTGAATTTATATACACA 3313
QY      788 oCysIleProGlnGluThrSerIleIleLeuPheAspLeuSerGlyLysIleValMetLys 808
Db      3314 AGAATTTGAAGAACTATCAATCACTAAAGAGTGCATTTAGACTAAAT---GTATGCG 3370
QY      808 sAsnSerLeuSerAlaGlyHisGlyArgMetAspValSerArgLeuProAsnGlyAlaTyr 828
Db      3371 TTATGATGCGTGAATGTGCTGGAATACAAATGATGAACGTTTGAATAAAATACCGACTT 3430
QY      828 rIleLeuLysValAspGlyTyrThrThrLysIleAsn 840
Db      3431 C-----CAAGTAGATGTTTATGAAACAGATCTAAT 3461

```

```

RESULT 5
US-10-114-170-39
: Sequence 39, Application US/10114170
: Publication No. US20030023075A1
GENERAL INFORMATION:
APPLICANT: Blatter, Frederick R.
            Burland, Valerie
            Perna, Nicole T.
            Plunkett, Guy
            Welch, Rod
TITLE OF INVENTION: NO. US20030023075A1 Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESS: Quarles & Brady
STREET: 1 South Plinkney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/10/114,170
  FILING DATE: 01-Apr-2002
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 09/453,702
  FILING DATE: 03-DEC-1999
  APPLICATION NUMBER: 60/110,955
  FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
  NAME: Seay, Nicholas J.
  REGISTRATION NUMBER: 27386
  REFERENCE/DOCKET NUMBER: 960296,95017
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (608) 251-5000
  TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
  LENGTH: 25165
  TYPE: nucleic acid
  STRANDEDNESS: double
  TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-10-114-170-39

Alignment Scores:
Pred. No.: 0.00135      Length: 25165
Score: 145.00          Matches: 193
Percent Similarity: 34.75%      Conservative: 125
Best Local Similarity: 21.09%    Mismatches: 368
Query Match: 3.27%             Indels: 230
                                Gaps: 48
US-10-030-330-1 (1-843) x US-10-114-170-39 (1-25165)
QY 67 GlyIlethrsErGlnGluGlySerProAlaTyrPheTyrValAlaAsnArg-GlyAs 86
DB 2787 GCGGTAGCGGAGGAAATCAAGCGCGTGTGTCAGTGGAAAGACGACAGCAATAC 2846
QY 86 nAsnGluTyrTyrAlaLeuValAlaAlaAspAspArgIleProThrIleLeuAlaTyrSe 106
DB 2847 GGTCAACGGCGCGACAGTACTAAGCGTGCATACCCCAACGCCAAGC---CTTGCCTACA 2903
QY 106 rProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMetTyrLeuGlnI 126
DB 2904 CACCATCGCTCAG-----GACACACATT-----ATCAGTGC 2933
QY 126 eTyrAspGlnGluIleGlyLeuIleLeuSerGly-----LysAlaGlnLeuAsnGlu 144
DB 2934 TCGCGAACATTAACGTCGCGCTGTACTGAGCGGACGTCGAAATGAGAAAGCGGGCAAC 2993
QY 144 uIleLeuArgThrGlnGluValProAlaGluValAlaAlaLeuMetAspAsnGlyHisPh 164
DB 2994 CTTAACATGACCGCTCAACGGGAAAGCCATACAGAACCCGTCGAGTGGAGCGACGA--- 3048
QY 164 eAlaAsnAspPrometArgTyrAsnGlnGlyTyrProTyrAsnAsnLysGluProLeu 184
DB 3049 -----ACCTGGCAAGTGACGCTGCTGCCACGGAAGTCCAG---GCACT 3089
QY 184 uProAsnGlyAsnHisAlaTyrThrGlyCysValAla-----ThrA 198
DB 3090 GCGCGAGGGTAATTACGCTGTCAATGACGATGACAGCATCGGCGAGGGAACACCAACG 3149
QY 198 lAlaAlaGlnIleMetArg-----TyrHisSerTyrProLeuGlnGlyGluGlyS 215
DB 3150 CCACAGCGGGAATTTCATGAGTACACACCTCAGACCCCGTGC----- 3190
QY 215 ePheAspTyrHisAlaGlySerLeuValGlyAsnTyrSerGlyThrPheGlyGluMet 235

```

```

DB 3191 -----TCAAGTGTAAATACCGTGGCGGACGATA----- 3220
QY 235 yAspTyrPheAsnMetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValA 255
DB 3221 -----TTCTTAATATATGCGAGCAGCGCCGCG-----CCGAGATCATTCGCCG 3263
QY 255 sPaLaTyrAlaThrLeuMetArgAspValSerAlaSerValSerMetSerPheTyr---- 273
DB 3264 ACAAGTCAGCGGCTTCCTCCAGCGCATACG---GTACGGTGAATATGGGCACTCATGTCC 3322
QY 273 ----- 273
DB 3323 TGACGGGCATCGCTGCGCAGATGCGAAGCTGGAATGTGGCGCTGGACCCAGCGGTAAACC 3382
QY 274 -----GluAsnGlySerGlyThrTyrSerValTyrValGlyAlaLeuArgAsnA 291
DB 3383 GCACGCTGATCGCGAGCCCAATACGATTTTCTCACCGAGATGCTGAGAAATA 3442
QY 291 snPheArgTyrLysArgSerLeuGlnLeu---HisValArgAlaLeuTyrThr-----S 308
DB 3443 CTGGCGCGCGCTCGACGATCAACGCTGGTGGTCTCTCCGTGATCACCATTAAACA 3502
QY 308 eArgGlnGluTyrHisAspMetIleArgGlyGluLeuAlaSerGlyArgProValTyrTyrA 328
DB 3503 CCCTTCGCCGCGATGACATTATCATGTGC---GCAGAAAAGGTGGCGCCACTGACCTTA 3559
QY 328 lAglyAsnAsnGlnSer----- 337
DB 3560 CCGGTAGCACTCAACAGGCTGAGACAGCAAGCAACCGTCACAGTAACCTGCTGACAGA 3619
QY 337 lAheValCysAspGlyTyrAlaSerAspGlyThrPheHisPheAsn-----TyrG 354
DB 3620 GTTTTACCACTACCGTGCAGCGC---GATGGCTCTGGAGTGTGACCGTACCTGCCCGC 3676
QY 354 lTyrPheGlyGlyLysSerAsnGlyPheTyrLysLeuThrLeu---LeuSerProThrSer 373
DB 3677 CGATGGGAATATCGCTGACGCGCGCGTGGCGATTAACCGTTCTGTGACGATCTACGCG 3736
QY 373 eGluGlyIleGlyGlyGluGlyIleGlyPheThrIle---TyrGlnGluIleIleThrGly 392
DB 3737 GCATATCCCGGCAACACTTCCCGCACCATTCACCTCGATACGACGCGCCGCTTAAGCA 3796
QY 392 lGluProAlaLys-----ThrProAlaGluAlaGlyThrAspAlaLeuP 407
DB 3797 TTGATCCACTGACCGCTGATACATCATTAACCGCGGAAACGGGACAGAT---CTGC 3853
QY 407 rIleLeuAlaLeuLysAspIleGlyAlaGluTyrLysSerGluSerGlyLeuAsn---- 425
DB 3854 CCATACACCGGCAACACCGACGCTACCGCGGGGACAGACGTCGACCGTTACCTTAATGGC 3913
QY 426 -----ValGlyTyrSerIleTyrA 432
DB 3914 AGACGTATCAGGCGCTGCTGACAGCAGACGCACTGAGACGTCGACCGCGCCCA 3973
QY 432 snThrGlyGluGlnGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyG 452
DB 3974 ACCTGGGCGCA-----CTGGCTGACGCA 3997
QY 452 lValIleGlyValLysThrSerSerIleAsnIleSerTyrGlyTyrGlyGluHisP 472
DB 3998 ACAGCT-----ACGGTACCGCCACGCTGAAC-----GATGTCCCGGTAATC 4039
QY 472 rGluSerPheSer-----LeuAlaP 479
DB 4040 CGAGCAGCGTTTACACCGTGGCGCTGTGGATGCCACCGCGGTGTAAACATTAAATC 4099
QY 479 rAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuLeuTyrArgArg-----Thc 497
DB 4100 CGGTGGCAGACCGTAACGTCATCAACACGCGGGAACATGCTCAGCGCAATCATCAGCG 4159
QY 497 lThrGluGlnTyrGluProValArgHisAlaGlnGlyGlyTyrValAsnSerIleLys 517

```

```

Db      4160 GCACG-----GTTACTGGCCCTCAGCGCGCATATCTCACCCTGACGC 4204
QY      517 aLaenThrThrAspProAsnAsnValValValThrValAlaAsnAsnGlnGlyLeuS 537
      : : : : :
Db      4205 TGAATATATGTGAT-----TACACACAGCGGTGTGATGTTCGGCAACTGGA 4252
QY      537 erlle---ValProAsnSerPheValAlaAsnLeu-----AsnSerTyrGlnHisSer 553
      : : : : :
Db      4253 GTCTGGCGGCTTCGGCTCGGTGTCAGTGGCTGGCGGACGGCATTTATCTCTGACGC 4312
QY      554 -----ThrIleThrValGlnPheAsn 561
      : : : : :
Db      4313 TCTCGGTACCGCAACGCAAGCGGAAACAGCGGACCGCATCTGACCGTCAATA 4372
QY      561 eTAspSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrGlyAlaThra 581
      : : : : :
Db      4373 CCGCGCGGCC-----CTTATCCGCGATTACACGATTGCG---G 4408
      : : : : :
QY      581 LaAspAspValIleSerLeuGlyTyrValMetAlaGluValPro---GlyGlySerSerA 600
      : : : : :
      : : : : :
QY      4409 GCGATGATGTATTAAACCCAGCAAAAGGCGCGATCTCCAGATTACCGGACCGCAGC 4468
      : : : : :
      : : : : :
QY      600 snTyrProValValTyrSerLysAspValLeuThrLeuSerGlnGlyAspTyrThrLeuT 620
      : : : : :
      : : : : :
Db      4469 ATCAGCCGTGTACACCGCCATCAGCTG---ACGCGAAGCGGCAAAATTACAC--- 4521
QY      620 rPtyrArgPheSerIleAsnAsnGlnLysAspGluTyrLysIleGlySerValServ 640
      : : : : :
      : : : : :
Db      4522 -----ACACAGACCGCGCTCCGCGCAACTG-----ACGC 4552
QY      640 allysThrProThrGlnGlyTyrThrHisProLeuPheGlnValGlyHisAsnGlnThrSerT 660
      : : : : :
      : : : : :
Db      4553 TCACCGTCCGCGATCGCGGTACAGCATTTAGGCGACGCAACTATACGGTAACGGCGG 4612
QY      660 hTyrThrLeuAspMet-----AlaHisAsnArgValLeuProAspPheT 675
      : : : : :
      : : : : :
Db      4613 CGGTGACACGAGATATCGGCAACAGCGCCACTGCGCATTAACGCTGCTGCGACAGCG 4672
QY      675 hTleuLysAsnLeuGlyLeu---ProPheAsnGlnGlyLeuValValValPheArgGln 693
      : : : : :
      : : : : :
Db      4673 CGCTGCCGCGGTGACCATTAATCCGCTGGCAACGCGAGATATTATTAACCGCCCGCG 4732
QY      694 -----ThrGlnSerSerSerGlySerLeuTyrPheAlaGln-----GluThrValH 709
      : : : : :
      : : : : :
Db      4733 CGCGCGTGGCGCAACCATCAGCGCGGAGGCTGCGCGGAGATGCGGACACGGTAA 4792
QY      709 IsIleLysGlnGlyGluThrPheValTyrLysProVal-----Y 722
      : : : : :
      : : : : :
Db      4793 CTATTACCTGGGTGTGAATACTTATACGGCGAGCGGTGGCGACAAATCTCACTGAGCGC 4852
QY      722 aLGIuGlyProIleProAsp-----GlySerTyrArgAlaThrLeuHisAlaP 738
      : : : : :
      : : : : :
Db      4853 TGGACGTTCCAGCGGCAATATTCAGCGCGCTGGGAAATGCGCATTTTAACTGAATGCT 4912
QY      738 heValAsnGlnGlnGlnLeuTyrLeuLysGlnGlyLysArgAsnTyrThrValLysIle 757
      : : : : :
      : : : : :
Db      4913 CAGTCACCAATCAAAACGCGACACCGGAGCGGCGGAGATATCAACCATCGACGCCA 4972
QY      758 --ValAsnGlnThrAlaValAlaGluLysSerSerGlnGluIleArgValPheProA 777
      : : : : :
      : : : : :
Db      4973 ATCTGCCGCGCTGGCGGTGCGATACGGCGGCGGATGCTGCTCAATATCATGACG 5032
QY      777 snProAlaArgAspTyrValGluIleSerAlaProCysIleProGlnIuThrSerIleT 797
      : : : : :
      : : : : :
Db      5033 ACGGCAGCGCGCTGTGTGTCACCGGACAGCTGGGGCTGCTGAAAGCACGCGCGCTTA 5092
QY      797 LeLeu-----PheAspLeuSerGlyLysIleValMetLysAsnSerLeuSerA 813
      : : : : :
      : : : : :
Db      5093 CGGTACGATTAAATATGTGGAATACACACTGGCGTGCAAGCGCATGATGAGCGAGCGG 5152
QY      813 laGlyHisGlyArgMetAspValSerArgLeuProAsnGlnAlaTyrIleLeuLysValA 833
      : : : : :
      : : : : :
Db      5153 TGGCGGTACGCGCGCGCGAGGTATGCGCGCTGGCGTGGCGGAGCGTTAAATATGCGGTTT 5212

```

```

QY      833 spGlyTyrThrThr-----LysIleAsnIleValHis 843
Db      5213 CAGCGGCAAAAGTACGCGCGGAAACTGCGTGCATTAACGAT 5253

RESULT 6
US-08-837-459-21
Sequence 21, Application US/08837459
Patent No. US20020006407A1
GENERAL INFORMATION:
APPLICANT: McKee, Marian L.
APPLICANT: O'Brien, Allison D.
APPLICANT: Machiel, Marian R.
TITLE OF INVENTION: Histidine-Tagged Intimin And Methods Of
TITLE OF INVENTION: Using Intimin To Stimulate An Immune Response And As An
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSER: Flanagan, Henderson, Farbow, Garrett &
ADDRESSER: Dunner, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/837/459
FILING DATE: 18-APR-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Fordis, Jean B.
REGISTRATION NUMBER: 32,984
REFERENCE/DOCKET NUMBER: 04995, 0023-00000
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 3106 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-837-459-21

Alignment Scores:
Pred. No.: 0.000188 Length: 3106
Score: 140.00 Matches: 167
Percent Similarity: 30.67% Conservative: 128
Best Local Similarity: 17.36% Mismatches: 331
Query Match: 3.15% Indels: 336
Gaps: 45

US-10-030-330-1 (1-843) x US-08-837-459-21 (1-3106)
QY      74 GlySerProAlaTyrPheTyrValAlaAsnArg-----GlyAsnAsnGlnGlyTyrAla 91
      : : : : :
Db      144 GGTTCGATGCTGTTTATTTATGTTATCATCAATTCATTGCAATTCGTAATGTAATTTATTTT 203
QY      92 LeuValAlaAlaAspAspArgIleProThrIleLeuAlaTyrSerProIleGlyArgPhe 111
      : : : : :
      : : : : :
Db      204 AATTCGGTTCGATTCAAACTGTAACTCATGATAGCTATCAG-----AATCGCCTT 257
QY      112 -----AspMetAspSerMetProAsp---Asn 119
      : : : : :
      : : : : :
Db      258 TTTTATACGTTGAAAACTGTGAACATGTTGCCGATCTTTCTAAATCGCAAGATATTAAAT 317
QY      120 Leu---ArgMetIrp-----LeuGlnIleTyrAspGlnIuIleGlyLeuIleLeu 135

```



```

OY 726 IleProAspGlySerTyrArg---AlaThrLeuHisAlaPheValAsnGlyGlnGln 744
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2139 GTTAAATATCATCCGTTATCATCTTCACAACTTGGATGTCACAGCTAAAGTCTCA 2198
OY 745 LeuTyrLeu-----Lys 748
Db 2199 ACGCAAGCAACCCGGAATGATGTCGCGACGATPACACTCACTCCAGTCCGCC 2258
OY 749 GlyTyrArgAsnTyrThrValLysIleValAsnGlyThrAlaValGlnAlaIleGlnSer 768
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2259 GGTAAAGCACTGTTAGTCCGACAGTCACTGATGGGCTGAGGTTAAAGCGACTGAGTCTC 2318
OY 769 Ser-----GlnGlnIleArgVal----- 774
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2319 ACTTTTGTGATGAACTGAAATATGACAAAGTTGATATTTGTTACATGTCACA 2378
OY 775 -----PheProAsnProAlaArgAspTyrValGlnIleSerAlaProCys----- 789
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2379 GCGGAGTTCCTAAATATTTGGCTGCATATGCTTAAACTGAAAGCAAGCGGTGCT 2438
OY 790 -----IleProGlnGluThrSerIleIleLeuPheAspLeuSerGly 803
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2439 GATGCTACATATTCATGCTATTCAGAAATACCAAGATCGCAGCTGTCATCATCAGCG 2498
OY 804 LysIleValMetLysAsnSerLeuSerAlaGlyHisGlyArgMetAspValSerArgLeu 823
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2499 AAGTCACCTTGCAT-----GGTAAAGCGACGTCTC---GTATTTAAAGCC 2540
OY 824 ProAsnGly-----AlaTyrIleLeuLysValAspGlyTyrThrThrLys 838
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2541 ACATGTGGATGATACCAACAGTAACTTAACTATMAAGCACCGCTCGTATATGATMAAA 2600
OY 839 IleAsn 840
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2601 GTGGAT 2606

```

RESULT 7
 US-10-114-170-18
 ; Sequence 18, Application US/10114170
 ; Publication No. US20030023075A1
 ; GENERAL INFORMATION:
 APPLICANT: Blattner, Frederick R.
 Burland, Valerie
 Perna, Nicole T.
 Plunkett, Guy
 Welch, Rod
 TITLE OF INVENTION: No. US20030023075A1el Sequences of E. coli O157
 NUMBER OF SEQUENCES: 265
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Quarles & Brady
 STREET: 1 South Plinckney Street
 CITY: Madison
 STATE: WI
 COUNTRY: US
 ZIP: 53701-2113
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Word Perfect 8.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/114,170
 FILING DATE: 01-Apr-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 09/453,702
 FILING DATE: 03-DEC-1999
 APPLICATION NUMBER: 60/110,955
 FILING DATE: 04-DEC-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Seay, Nicholas J.
 REGISTRATION NUMBER: 27386

```

REFERENCE/DOCKET NUMBER: 960296, 95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 3695
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-10-114-170-18

Alignment Scores:
pred. No.: 0.000277 Length: 3695
Score: 139.50 Matches: 211
Percent Similarity: 31.32% Conservative: 111
Best Local Similarity: 20.53% Mismatches: 372
Query Match: 3.14% Indels: 336
DB: 9 Gaps: 47

US-10-030-330-1 (1-843) x US-10-114-170-18 (1-3695).
OY 63 AspaIaGluArgGlyIleThrSerGlnGlnGlySerProAlaTyrPheTyrValAla 82
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 526 GACGGTCAGAGCCGATTAATGCTGCTTAACGGTAATGAGCAAAACCGCTGCTG 585
OY 83 AsnArgGlyAsnAsnGlnGlyTyrAlaLeuValAlaIleAspAspArgIleProThrIle 102
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 586 TCTTGGCCGACGCGCGGCGACGACGTACAGCGGCGATGAAAGATCATGATCAAGACTAA 645
OY 103 LeuAlaTyrSerProIleGlyArgPhe-----AspMetAspSerMetProAspAsn 119
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 646 CTAACCTTCAAAACCGCGTGAATATTTGTAAGTCTGCTCCCTGAAGCCACATAACACAG 705
OY 120 LeuArgMetTyrLeu---GlnIleTyrAspGlnGlnIleGlyLeu----- 133
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 706 GCMAAGCCCAACACTGGTGATTCACCGAACTGAAGCAGGCGGTATATCAGTCTGCTTT 765
OY 134 -----IleLeuSerGlyLysAlaGlnLeuAsnGlnGlnIle---LeuArgThrGlu 149
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 766 ACTACCGGNAAGCAGTCAAGTGAAGCAACGATTACTGTTAGCGTTGATGACATGACAG 825
OY 150 GlyValProAlaGlnValHisAlaLeuMetAspAsnGlnHisPheAlaAsnAspPromet 169
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 826 ACCGTCACTGACGAACCTGCGGCCACAGTATGAT-----GTGCAAACTCCACCCCTG 879
OY 170 ArgTyrAsnGlnGlyTyrProTyrAsnAsnLysGluPro-----LeuLeuProAsn 186
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 880 AGCGCTAAC-----GAGCCGTACGTGACGTGCTGCTGAT 915
OY 187 GlyAsnHisAlaTyrThr-----GlyCysValAlaIle 197
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 916 GGTACGACAAAGCTATATGACGTGACGTGCGGTGACCTCCGAGGTAATCCGGTGACG 975
OY 198 AlaAlaValaGlnIleMetArgTyrHisSer-----Tyr----- 208
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 976 GGAGAAGCAGCCCGCTTGCGATTTGT-TCGGAACACACTAATGTTGAACCGTTGGTGC 1034
OY 209 -----ProLeuGlnGlnGlySerPheAsp----- 217
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1035 CATTCGGAATATAAACCAGGCGTTTACAGCGCCCGCTTCTTGACCCGTCGCGGANA 1094
OY 218 -----TyrHisAlaGlySerLeu----- 223
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1095 CGTGTGTGCGTCTTTCAGCGACAGTATCAGCTGGGACATTACAAACAACCGCTGAA 1154
OY 224 -----Val 224
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1155 GTTTCGCGCGGCGCTTGTATGACGACATTCGTCATCACCCCTGAATCCTGATTAAC 1214
OY 225 GlyAsnTyrSerGlyThrPheGlyLysMetTyrAspTyrIleAsnMetProGlyAsnPro 244

```

```

Db      1215 GGTGGTGGGGGACAGTACGGC-----AATCGACGGTAAGA 1256
Qy      245 Aspleu-AspAsn-----LeuThrGlnSerGlnValAspAlaTyrAlaThrLeuMetArgAs 263
Db      1257 TGCCATGACCAACCCCTGTGACAGCCCTCAAGCCGCGCTCATTTAGCGGGTCCGC 1316
Qy      263 pValSerAlaSerValSerMetSerPheTyrGlnAsnGlnSerGlnTyrSerValTy 283
Db      1317 TGCTAAGAGTTCTACGGCATCGGGCTGCACAAATATATGGTATGGAGCTGGACCTGC 1376
Qy      283 rValValAlaLeuArgAsn-----AsnPheArgTyrLysArgSer----- 297
Db      1377 GATTACTCTCGGCTCTACGGGGGCGGTGAATTAGAGTTATGCCAAGCTTAATGACAGAA 1436
Qy      298 -----LeuGlnLeuHis-----ValArgAlaLeuTyrThrSerG 309
Db      1437 TGCGGCACCAATGCGGGCAAAAGTAACCGGTGGCTGATGCGTTATCTTCAAAACAGTC 1496
Qy      309 InGU-----T 311
Db      1497 GAAATCTCTGTGCGAAGATACAGTAAACCGGCGCAAGACACACCGTGACGCTGCT 1556
Qy      311 rPHisAspMetIleArgGlnLeuAlaSerGlnYArgProValTyrTyrAlaGlnAsnA 331
Db      1557 GCGCAAAAGATGCG-CATGGCAACGCTATCATGCTG----- 1590
Qy      331 snGlnSerIleGlnHisAlaPheValCysAspGlnTyrAlaSerAspGlnThrPheHisP 351
Db      1591 -----CTTGCCTGTGCGCAAGTTGACGCGGCGCCCTCTTCAGAGGCCAGCAGCCGTTT 1642
Qy      351 heAsnTrpGlnTyrPheGlnValSerAsnGlnPheTyrLysLeuThrLeuLeuSerProT 371
Db      1643 CCAGTTGGACCGCAAAAGGT-----AACGTTCTATGTGTCTACGTTGACTACAGGTG 1696
Qy      371 hrSerLeuGly----- 374
Db      1697 GAAACAGCGGCGAGCTGCGCTCATGCTCTCTTCAGAGGCCAGCAGCAGCAGCAGAG 1756
Qy      375 -----IleGlnGlnGlu----- 378
Db      1757 CCGCGCAGTTGACGGGTATCGCGGAGAGATGTCATACAGCAACTCTACGCTTGTTCGG 1816
Qy      379 -----GlyIleGlnPheThrIle-----T 385
Db      1817 ACAATRAGGCTCCGACCGTCAAACAGCAGCAGCAACTCACCTTCACGTCGAAGAGATGCT 1876
Qy      385 yrgInGUleIleIleThrGlnIleGluPro-----AlaLysT 397
Db      1877 ACGGGAACCCGGGTACCGGGGTGAAGCCAGATGCACCACTGTTTAGCGGTGGCCGACGA 1936
Qy      397 hrProAlaGlnAlaLeuThrAspAlaLeuProIleLeuAlaLeuLysAspIleGlnAlaG 417
Db      1937 CGGGAGTGAGGCGCTTCA-----GCAG 1960
Qy      417 LuTyrLysSerGlnSerGlnLeuAsnValGlnTyrSerIleTyrAsnThrGlnGlnG 437
Db      1961 GAAATGACAGCAAGAAAGTAAATGAGGTCTACGTGTGACCTTAACGCTGGAGTCTGCCG 2020
Qy      437 InSerAsnLeuAspLeuGlnTyrArgLeuAsn----- 447
Db      2021 CGGGTACGTTGCTGTGATGCGCGAGTGAACGGCAAAATGCCGTGCTCAGCCATCGG 2080
Qy      448 -----LysAlaAspGlnGlnValIleGlnValLysThrSerIleAsnIleS 464
Db      2081 TGCTGAACGTTGACAGTGCAGCATTAAGGCTGAGATTCGTATATGACAGTGAAGGTAA 2140
Qy      464 eTrpTyrGlnTyrGlnHisProGlnSerPheSerLeuAlaProAsnGlnLeuSerG 484
Db      2141 AT-----AACCAACTGGCTA 2155
Qy      484 InGU-----IleAsnThrIleThrLeuLeuTyrArgArgThrGlnGlnIntProG 502

```

```

Db      2156 ATGACAGCTGTGTAACCAAGATTAACCTG----- 2184
Qy      502 LuProValArgHisAlaGlnGlnGlnTyrTyrValAsnSerIleLysValAsnThrAsp 522
Db      2185 --ACCGTTGTGACACCTATGTAACCCGTGGAGGGCAGAGAAATTACGCTTATAC 2242
Qy      522 roAsnAsnValVal-----ValThrValAspAsnAsnGlnGlnLysLeuS 537
Db      2243 CGCAGGCTGTGACACCAAGCAGCGGGAATACATAACCACTAATTCGGCGCAATTAACCGG 2302
Qy      537 eTrIleValProAsnSerPheValAlaAspLeuAsnSerTyrGlnHisSer----- 553
Db      2303 ACATTAGAGTTATGTCACAGGTTGCCGGAACACACATATTTCCGCTGCGTGAATAGTG 2362
Qy      554 -----ThrIleThrValGlnPheAsnSerAsp--SerProAspGlnIleArgThrP 570
Db      2363 CTCAGAAAGACGCGACGGGTGAATTCACCGGATCCACACCGGTCAGCAAGAACTGTC 2422
Qy      570 roValAlaPheAlaLeuSerThrGlnValAlaThrAlaAspAspValIleSerLeuGlnTyrPv 590
Db      2423 AGGTACAGCCCGCTGCTCAAAAGTGCCAAACGCGCAAAAGATGCTTACCTGACCGCGCA 2482
Qy      590 aLMetAlaGlu-----ValProGlnGly-----SerSerAsnTyrPro- 602
Db      2483 ACGTGAGATTAATAATGTAACTTTCAGGAGACCTGTGACCTTATATCTGCCCC 2542
Qy      603 -----ValValThrSerLys----- 607
Db      2543 GGGGTGTCAAGCCGCTTACAGCGGATATCTGTGGGTGAAGCAACGATGAGGGGAAG 2602
Qy      608 -----AspValLeuThrLeuSerGlnGlnLysAspTyrThrLeuTyrArgPheSerI 625
Db      2603 CAGAGTTGACAGGTGTTTACGTACTGCGCGAAGCTGTGATGATCAGCGCATCGGCAGCG- 2661
Qy      625 leAsnAsnGlnLysAspGlnTyrPheLysIleGlnSerValSerValLysThrProThrG- 645
Db      2662 --AATAGCCAGCCTTGGAATACGACGACATTAACGTTTGAAGCGAATGCTACCGCAA 2719
Qy      645 LuTyrThr-----HisProLeuPheGlnValGlnHisAsnGlnT 658
Db      2720 CCGTTCGCCGTATTTAGGTGATGTCGCAACTATCTACCTGCGGCGCAATGCCAACAAGA 2779
Qy      658 hrSerThrTyrThrLeuAspMetAlaHisAsnArgValLeuProAsp-----PheThrL 676
Db      2780 CGTATAAAGTTACGGTGACTGATGCCAATATACCACTGTTGAAGATAGCGAAGTGAAGC 2839
Qy      676 eu-----LysAsnLeuGlnLeuProPheAsnGlnGlnLeuValValPhea 692
Db      2840 TGACTGCCAGCCCGCAAAATTTAGTCTGCTCCCATGGGAGC-----G 2884
Qy      692 rGlnThrGlnSerSerSerGlnSerLeuThrAlaAlaGlnGlnGlnThrValHisIleLysG 712
Db      2885 CGAAATCATATGACAGCAAGACAGCTATTTTCCCGCCACAGCACTGTGTCAGCGA- 2943
Qy      712 InGUleGlnThrPheValTyrLysProValValGlnGlnProIleProAspGln-----S 730
Db      2944 -----TATACSTCAAGCGCGCAAAAGTGAAGTACGCGCGAGCTGCAGAGAT 2986
Qy      730 eTrTyrArgAlaThrLeuHisAlaPheValAsnGlnGlnGlnGlnLeuTyrLeuLysGlnL 750
Db      2987 CGAGCAAAACSTGCGCAATCTAATTTGCTGCGGATGATACAAATGCAAGTACTACCGCAT 3046
Qy      750 ysrArgAsnTyrThrValLysIleValAsnGlnTyrThrAlaValGlnAlaIleGlnSerSerc 770
Db      3047 CATCTGATGTGATCTCTGTGTCGCGAGTGGATATACGACTGCGAAGCTGAGAGTGAAC 3106
Qy      770 InGUleArgValPheProAsnProAlaArgAspTyrValGlnIleSerAlaPro----- 788
Db      3107 TGATGTGCGCAATTAACCCCGTTGGGGGGAATATGTGGGTGCACATTAAGCCCGCAAG 3166
Qy      789 -----CysIleProGlnGlnThrSerIleIleLeuPheAspL 801
Db      3167 GGTGACGAGAAAGATTAATCACTTCTGCGCGCAAAATGACCATTTCC-----G 3217
```



```

Db      1658 GCAGGCGTTTCGTGGAGAGAAAGTACTGTTGATGGCAGGTGCATATCCAGAGGC 1717
      475 PheSerLeuAlaPro----- 479
Db      1718 GATCGCTTATCCAAATTCGGTGGACCTTACTGTTAATGCTAAAGGAAACTTA 1777
      480 AsnGlnLeuSerGlnGlyLeuSerThrIleThrLeuLeuTyrArgArgThrGlyThrGlu 499
      1778 GGAAGCCTGAGTGGGTACGAGTGGTGGTGTGA-----GATCAACAGCAGATGGA 1831
      500 GlnTProGluProValArgHisAlaGlnGlyTyrValAsn-----SerIleLys 516
      1832 TCAGGTCAAAAACCAAGCCCTTAAAGAGTGGCATTTGAGTGTAGCTACGCTACCTCA 1891
      517 ValAsnThr-----ThrAspProAsnAsnValVal-----Thr 528
      1892 CTAATATAGTCAGATCAAGTGTGATCTTACCAATATTATTGCGCTTCGTGGTGGCC 1951
      529 ValAspAsnAsnGlnGlyLysLeuSerIle-----ValProAsn-----Ser 542
      1952 TTAGATCTTAATGGGCTTCAATTAACCTTTGAACGATCCAAATACGATGAAGCGCG 2011
      543 PheValAlaAspLeuAsnSerTyrGlnHisSerThrIleThrVal-----Gln 558
      2012 ATGATTTGTCACACACAGCTCTCTCAACCGCAATATATACGATTACAGCAACGCACT 2071
      559 PheAsnSerAspSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrGly 578
      2072 ATTAATTCAGATAGCAACACTT----- 2095
      579 AlaThrAlaAspAspValIleSerLeuGlyTyrVal-----MetAlaGluVal 594
      2096 ACTAATTAATAAAAGATTCATTTAAGCGCTGCTGGTGGTGGAGCAGATTAAGCTAAACA 2155
      595 ProGlyGly---SerSerAsnTyrProValIleTyrSerLysAspValLeuThrLeuSer 613
      2156 AATGCTGCTTAATGTAATGTAATATCAACGATTAATGCAAGAAATCATTTGCTGCTCT 2215
      614 GlnGlyAspTyrThrLeuTyrPyrArgPheSerIleAsnAsnGlnLysAspGluTyrPlys 633
      2216 GGGGG-----ACAAATTTAAACGCAATATACAG 2245
      634 LysIleGlySerValSerVal-----LysThrProThrGlnIleTyrThrHisPro 649
      2246 CAAATGCGTGTACCTTACTTTAGTGTGCTCCAGCCCTCATGCTTACATCAT--- 2302
      650 LeuPheGluValGlyHisAsnGlnThrSerThrTyrThrLeuAspMetAlaHisAsnArg 669
      2303 -----TTAAGAAGAGACTTGTCTAACATGAA 2329
      670 ValLeuPro-----AspPheThrLeuLysAsn 678
      2330 GGTATCCCAAGGCAAGGAAATGTTGTGGATCAGATTCGATCAACCGCATTTAAAGCT 2389
      679 LeuGlyLeuProPheAsnGlnGluLeuValValValPheArgGlnThrGlnInserser 698
      2390 GAAAACCTCCAAATTAAGCGGAGAGTGGGTGCTTCGCAATGTTTCTTCAATTTGAG 2449
      699 GlySerLeuTyrAlaAlaGlnGlnIleValHisIleLysGlnGlyLysThrPheValTyr 718
      2450 GGAAT---TGG-----ACAGCAGCAATATGCAAAATCCCATTT----- 2488
      719 LysProValValGlnGlyProIleProAsp 728
      2489 -----GGTGTGTGCCAAAT 2503

```

```

TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
FILE REFERENCE: A-59941-1/RT/DCF/DHR
CURRENT APPLICATION NUMBER: US/10/080, 505
CURRENT FILING DATE: 2002-02-22
PRIOR APPLICATION NUMBER: US 08/296,791
PRIOR FILING DATE: 1994-10-25
PRIOR APPLICATION NUMBER: US 09/839,996
PRIOR FILING DATE: 2001-04-20
NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14
LENGTH: 4822
TYPE: DNA
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: CDS
LOCATION: (388)..(4563)
OTHER INFORMATION:
US-10-080-505-14

Alignment Scores:
Score: 0.00053 Length: 4822
Percent Similarity: 138.50 Matches: 165
Best Local Similarity: 31.368 Conservative: 89
Query Match: 20.378 Mismatches: 275
DB: 9 Indels: 281
Gaps: 44

US-10-030-330-1 (1-843) x US-10-080-505-14 (1-4822)
Db      67 GlyIleThrSerGlnGlnGlyLysSerProAlaTyrPhe-----TyrVal 81
      439 GGGATGTATCGACACCGTGGCAGCGTACACCTATTATTTGGGATGACTACCAATATTAT 498
      82 AlaAsnArgGlyAsnAsnGlnGlyTyrAlaLeuValAlaAlaAspAsp----- 97
      499 CGTGAATTTGCCAGATAAAGGAGATTACAGTTGGGGCTCAGATATGATATCTAC 558
      98 -----ArgIleProThrIleLeuAlaTyrSerProIleGlyArgPheAsp 112
      559 AATTAATAAAGGGAATGATAGTACGATGATGAAGTGTGCTATGCT-----GAT 612
      113 MetAspSerMet----- 116
      613 TTAATCTTCATGTTGCTGCTGCTATTCACATTTGATAGTACGACATTTAATAC 672
      117 -----ProAsp 118
      673 GTCCACATTAATGATGATGATGATGATGATTTGGTATGGAGGGGAAATCCAGAC 732
      119 AsnLeuArgMetIlePheGlnIle-----TyrAspGlnIleGlyLeuIleLeu 135
      733 CAACATCGTTTAAAGTATAAAGTTGTTAAACGATATATATTAAG----- 777
      136 SerGlyLysAlaGlnLeuAsnGln-----GluIleLeuArgThrGlu 149
      778 AGCGGTGATAGACAAATATATGATTTCAACATCCAGATTAAGAAATTTGAACGGA 837
      150 GlyValProAlaGluValHisAlaLeuMetAspAsnGlnLysIlePheAlaAsnAsp----- 167
      838 ACTGACACCTATTAATTAATGTTTCAATATGATGATGATATATCAAAATTTTAATCAA 897
      168 -----PrometArgTyrPheAsnGlnGlyTyrProTyr-----AsnAsnLys 180
      898 TATCTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 957
      181 GluProLeuLeuProAsnGlnLysAsnHisAlaTyrThrGlyCysValAlaThrAlaAla 200
      958 -----ACCATTTGAGACCTTAGCTTAGAGCT----- 964
      201 GlnIleMetArgTyrHisSerTyrProLeuGlnGlyLysSerPheAspTyrHisAla 220
      985 -----TCATGTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1026

```

RESULT 9
 US-10-080-505-14
 ; Sequence 14, Application US/10080505
 ; Publication No. US20030073166A1
 ; GENERAL INFORMATION:
 ; APPLICANT: St. Geme, Joseph W.

```

OY 221 GlySerLeuValAlaGlyAsnTrpSerGlyThrPheGlyGluMetTyrAspTrpIleAsnMet 240
DB 1027 GGTAACT-----
OY 241 ProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAlaSerAlaThrLeu 260
DB 1033 ---GGTAACTTAAATTAATGGCGAGTCAAAATCTT---AATAAATATGGTCCACTA 1086
OY 261 MetArgAspValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThrTyr 280
DB 1087 -----CCTACGGCAGCTTCATTCGGGATAGAGTTCCTCCATG--- 1125
OY 281 SerValTyrValAlaGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGlnLeu 300
DB 1126 -----TTTATTATGATAAAGAACTT----- 1146
OY 301 HisValArgAlaLeuTyrThrSerGlnGluTyrHisAspMetIleArgGlyLeuAla 320
DB 1147 -----AAGAAATGG-----TATTAATATGGCGTGTACGT 1176
OY 321 SerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCys 340
DB 1177 GAAGAAATCCTTATGCTGACAGTAGAAGAACAGTATCAATTAACGAAAAAGATTATTTT 1236
OY 341 AspGlyTyrAlaSer---AspGlyThrPheHisPheAsnTrpGlyTyrGlyValSer 359
DB 1237 CAAGCTATTCTTAATCAAGCATTAACGCTAATTTT---TGGGAT-----ACT 1281
OY 360 AsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGly 379
DB 1282 AATGCTGAATTAACATTT-----AATATAGGAGTCAACCAATGGA 1323
OY 380 IleGlyPheThrIleTyrGlnGluIle---IleThrGlyIleGluProAlaLysThrPro 398
DB 1324 AGAGTGCACAACAACTACATTAACCTAAAGAAAGCTATTCACGCTGAAGCATAGTG 1383
OY 399 AlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLeuLys---AspIleGluAla 416
DB 1384 GGTCTTATGATTAATATGACCACTCATGATGCTAGATTAATAATAGGCGATGAATCTGCC 1443
OY 417 GluTyrLysSerLeuSer-----GlyLeuAsnValGlyTyrSerIleTyrAsn 432
DB 1444 TCTTAATAAAGTCTTAATCATGATGCGCAGCATTAACATCAATGGAAGAAATTTACTTT 1503
OY 433 ThrGlyGluGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyGlu 452
DB 1504 GCGCATCAAGAAAGAAAGCACTTTAAACAATTGAAATATATATATCAAGTGCAGGTGGA 1563
OY 453 Val-----IleGluValLysThrSerSerIleAsnIleSerTrpTyrGly 467
DB 1564 TTGATTTTGAAGTAATTTTGTGTAAAGGCAATCAAAATATATTAATCTGGCAAGGT 1623
OY 468 TyrGly-----GluHisProGlnSer 474
DB 1624 GCAGGCGTTTCTGTGAGAAAGAAAGTACTGTTGAATGCGAGTGCATTAATCAAGAGGC 1683
OY 475 PheSerLeuAlaPro----- 479
DB 1684 GATCGCTTATCAAAATTTGGCGTGGAGACCTTACTGTTAATGCTAAAGGAAAAACTTA 1743
OY 480 AsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuLeuTyrArgArgThrGlyThrGlu 499
DB 1744 GGAAGCTAGTGTGCTAAGCGTTGGTGTGTTA---GATCAACAAGCAGATGAA 1797
OY 500 GlnTrpGluProValArgHisAlaGlnGlyTyrValAsn-----SerIleLys 516
DB 1798 TCAGGCAAAAAACAAGCCTTTAAAGAGTGGCATTTGAATGGTAGACCTACCGTTCAA 1857
OY 517 ValAsnThr-----ThrAspProAsnAsnValValVal-----Thr 528
DB 1858 CTAATATAGCAGATCAATGATGATCTTAACAATATTATTTATTCGCTTCTGCTGTCGC 1917

```

```

OY 529 ValAspAsnAsnGlnGlyLysLeuSerIle-----ValProAsn-----Ser 542
DB 1918 TTACATCTTAATGGGCAATTCATTAACCTTTGAAACGATCAACAAATACGAGTAAGCGCG 1977
OY 543 PheValAlaAspLeuAsnSerTyrGlnHisSerThrIleThrVal-----Gln 558
DB 1978 ATGATGTGAAACCAACACCTTCTCAACCCGCAAAATATATACGATTAACGACCAAGCACT 2037
OY 559 PheAsnSerPheSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrGly 578
DB 2038 ATTAATTCAGATTAACCAACACTT----- 2061
OY 579 AlaThrAlaAspAspValIleSerLeuGlyThrVal-----MetAlaGluVal 594
DB 2062 ACTAATTAATAAAGATATTCATTAACGCTGTGTGGTGAGCAAGATTAAGCTAAACAA 2121
OY 595 ProGlyGly---SerSerAsnTyrProValValIlePheSerLysAspValLeuThrLeuSer 613
DB 2122 AATGCTCGTTAAATGATGATATATCAACACAGTTAAATGACAGAAATCATTTGTTGCTTCT 2181
OY 614 GluGlyAspTyrThrLeuTyrArgPheSerIleAsnGlnLysAspGluTrpLys 633
DB 2182 GGGGGG-----ACAAATTTAAACGCAATATACAG 2211
OY 634 LysIleGlySerValSerVal-----LysThrProThrGluTyrThrHisPro 649
DB 2212 CAATAATGTGTACGTAGTTAGTTAGTGTGCTGCTCAACGCTCATGCTTACATATCAT 2268
OY 650 LeuPheGluValGlyHisAsnGlnThrSerThrTyrThrLeuAspMetAlaHisAsnArg 669
DB 2269 -----TTAAGAGAGCACTTCTCAACATGGA 2295
OY 670 ValLeuPro-----AspPheThrLeuLysAsn 678
DB 2296 GGTATCCCAACGCGCAATTTGTGGATCAAGATTGATGATCAACCCGACATTTAAAGCT 2355
OY 679 LeuGlyLeuProPheAsnGlnGlyLeuValValPheArgGlnThrGlnSerSer 698
DB 2356 GAAACTTCCAAAATTAAGCGGAGTGCAGGTGCTTCGCAATGTTCTTCAATTGAG 2415
OY 699 GlySerLeuTyrPalaIleGlnGluThrValHisIleLysGlnGlyGlnThrPheValTyr 718
DB 2416 GGAAT---TGG-----ACAATGCAATTAATCAATGACACATTT----- 2454
OY 719 LysProValValGlyGlyProIleProAsp 728
DB 2455 -----GGTGTGTGCCAAAT 2469

```

```

RESULT 10
US-10-160-758-4
Sequence 4, Application US/10160758
Publication No. US20030036076A1
GENERAL INFORMATION:
APPLICANT: EXELIXIS, INC.
TITLE OF INVENTION: CADS AS MODIFIERS OF THE P53 PATHWAY AND METHODS OF USE
FILE REFERENCE: EX02-089C
CURRENT APPLICATION NUMBER: US/10/160,758
PRIOR FILING DATE: 2002-06-03
PRIOR APPLICATION NUMBER: US 60/296,076
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/328,605
PRIOR FILING DATE: 2001-10-10
PRIOR APPLICATION NUMBER: US 60/357,253
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 10455
TYPE: DNA
ORGANISM: Homo sapiens
US-10-160-758-4

```

Alignment Scores:

Pred. No.: 0.00173 Length: 10455
 Score: 138.50 Matches: 199
 Percent Similarity: 33.90% Conservative: 127
 Best Local Similarity: 20.08% Mismatches: 336
 Query Match: 3.12% Indels: 329
 DB: 9 Gaps: 52

US-10-030-330-1 (1-843) x US-10-160-758-4 (1-10455)

QY 21 SerAlaProValThrLysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgGluVal 40
 DB 1420 AATGCCCGGAGTTCACAGCTCCGAGTACAGCTGGCCATCATCTGACAGCTGCACAGCTC 1479
 QY 41 SerLeuArgMetGlyGln---ThrAlaValSerAspLysIleSerIleAspLysValTyr 59
 DB 1480 GGCCTTGGCCCTCCACTCTTCACTCCAGGTGGTGACAG----- 1518
 QY 60 ArgGlnGlyAspAlaGluArgGlyIleThrSerGlnGlu-----GlySer 75
 DB 1519 -----GATGAGACCCAGGCGCTGAACAGCATGTTGAGGTACTTGGTGGGAGAC 1569
 QY 76 ProAlaTyrPheTyrValAlaAsnArgGlyAsnAsnGluGlyTyrAlaLeuValAlaAla 95
 DB 1570 AACTCCACACCACTTATCATCTCCCGACCTCCGACGAGGAGAGCGGACATTCCTATT 1629
 QY 96 AspaAspArgIleProThrIleLeuAlaTyrSerProIleGlyArgPheAspMetAsp--- 114
 DB 1630 CGGGTGGCCATCCCA-----CTGAGCTACGAGACCGGTGACCGCTTGTGATCTC 1683
 QY 115 -----SerMetProAspAsnLeu-----ArgMetLysLeuGlnIleTyr 127
 DB 1684 TTTCGCAATGAGAGGTGCTGACCATGTGGGTGAGGTGAGTGAAGATCATCTGCATC 1743
 QY 128 AspGlnGluIleGlyLeuIleLeuSerGlyLysAlaGlnLeuAsnGluIleLeuArg 147
 DB 1744 AATGAAATGACAAACCGGCCCTTCACAGCCACCTGTACAAATCATCCCTGTACAG 1803
 QY 148 ThrGlnGluValProAlaGluValHisAlaLeuMet-----AspAsnGly----- 162
 DB 1804 AACGTACACCGTGGGAGCTCTGTGACAGCTCCGACAGTGCATATGACAGCCACC 1863
 QY 163 -----HisPheAlaAsnAspProMetArgTyrPheAsnGlnGlyTyrPro 176
 DB 1864 TTTCGGAAGTACAGTACTTCTTCACTGATGACCTGACAGGTTCGCTG----- 1914
 QY 177 TrpAsnAsnLysGluProLeuLeuProAsnGlyAsnHisAlaTyrThrGlyCysValAla 196
 DB 1915 -----GACAAAGAC-----ACGGACTCATCATG 1938
 QY 197 ThrAlaAla-----AlaGlnIleMetArgTyrHisSerTyrProLeuGlnGly 212
 DB 1939 CTGATTTGCCAGGCTGGACTGATGAGCTCATCCAGCGCTTCACTGATGATGATGATG 1998
 QY 213 GluGlySerPheAspTyrHisAlaGlySerLeuValGlyAsnTyrSerGlyThrPheGly 232
 DB 1999 GACGGGGGGGGGAGAGACACACAGCGCGTCAAGATCAAT----- 2040
 QY 233 GluMetTyrAspTyrIleAsnMetProGlyAsnProAspLeuAspAsnLeuThrGlnSer 252
 DB 2041 ---GTGTGGAT---GTCAAC-----GACAACTGGCCCACTTC 2073
 QY 253 GluValAspAlaTyrAlaThrLeuMetArgAspValSerAlaSerValSerMet----- 270
 DB 2074 CAGAAAGATGCTTACGTGGTCTCTGCGGAGAAAGAGACCTTCTGTCTCACACAGCTGTG 2133
 QY 271 -----SerPheTyrGluAsnGlySerGlyThrTyrSerValTyr 283
 DB 2134 CGGCTCGGGGCAACAGATGAACATCCCTCCCAACACACATCATCATGACATTC 2193
 QY 284 ValValGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGlnLeuHisValArg 303
 DB 2194 AGTCATCTGCCCTTTGGAGAGCTACTTCGACATC----- 2226

QY 304 AlaLeuTyrThrSerGlnGluTyrPheHisAspMetIleArgGlyGluLeuAlaSerGlyArg 323
 DB 2227 AGCTTACAGAGGCTAT-----GGATGATACAGCTGACGTCC 2265
 QY 324 ProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCysAspGlyTyr 343
 DB 2266 CCGCTGCATTTATGAA----- 2280
 QY 344 AlaSerAspGlyThrPheHisPheAsnTyrPheGlyGlyValSerAsnGlyPheTyr 363
 DB 2281 -----CAGATATCAATGCGTCTGAT 2301
 QY 364 LysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGluGlyIleGlyPheThr 383
 DB 2302 TATCTACGCTATG-----GCCATGATGCTGGCAACCCCTCTCAACAGCAC 2352
 QY 384 IleTyrGlnGluIleIleThrGlyIleGluProAlaLysThrProAla----- 399
 DB 2353 GTCCGTGACCATCGAGGTGTTGATGAGATGACAAACCTCCACCTTCAGCAAGCCC 2412
 QY 400 -----GluAlaGlyThrAspAlaLeuProIleLeu 409
 DB 2413 GCCTACTGCTCTCGGTGTGAGAACATCATGACAGGACCCAGGTGCTTCTGTAAT 2472
 QY 410 AlaLeuLys---AspIleGluAlaGluTyrLysSerGlySerGlyLeuAsnValGlyTyr 428
 DB 2473 GCCACAGACCTGACCGCTCCGCGAGTACGGCCAGAGATCC-----ATCATCTAC 2523
 QY 429 SerIleTyrAsnThrGlyGlnGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLys 448
 DB 2524 TCCCTG-----GAAAGCTCACCCAG-----TTTCGATCAATATGCC 2559
 QY 449 AlaAspGlyGluValIle-----GluValLysThrSerSerIle 461
 DB 2560 CGCTCAGGGAAATACACACACACACCTCTGCTTACCAGGACCAACCTGTGAATATC 2619
 QY 462 AsnIleSerTyrPyr-----GlyTyrGlyGlnHisProLysPheSerLeuAla 478
 DB 2620 CTCTATGCTTCCGCACTGACAGGGGGGTGGCCACACCAAGAACT----- 2667
 QY 479 ProAsnGlnLeuSerGlnGlyIleAsnThr-----IleThrLeuLeuTyrArgArgThr 496
 DB 2668 -----GGCATGCCACCGTAACATCATCCCTCGACATCAATGAC 2709
 QY 497 GlyThrGlnGlnTyrPheLysProValArgHisAlaGlnGlyLysTyrValAsnSerIleLys 516
 DB 2710 AACCAACCCACCGTGAAGAGACGACCTTAC-----TACATCAACTGTGTGGAG 2757
 QY 517 ValAsnThrThrAspProAsn---AsnValValIleValIleAspAsnAsnGlnGlyLys 535
 DB 2758 ATGACCCCTTCACACTCTGATGTGACACCGGTGGCTGTGATCCACAGCTGGGGAG 2817
 QY 536 -----LeuSerIleValProAsnSerPheValAlaAspLeuAsnSerTyr 550
 DB 2818 AATGGACCTGTGTACAGCATCCAGCAACCAACAAAGTTTACAGCCTCAACAGACACC 2877
 QY 551 -----GluHisSerThrIleThrValGlnPheAsnSerAspProAsp----- 565
 DB 2878 ACGGGCAAGATCGCACACCCACGACCTGTGACCGGGAGAACCCGACCCCATGAG 2937
 QY 566 -----GluIleArgThrProValAlaPhe 573
 DB 2938 GCCGAGCTGATGCGCAAAATCGTCTGCTGTACTGACTGTGGAGGCCCTCTGAAA 2997
 QY 574 AlaLeuSerThrGlyAlaThrAlaAspAspValIleSerLeu----- 587
 DB 2998 GCCACACGAGATGCCACAGATGTTGTGAACCTTGTGATCAATGACATGACCCAC 3057
 QY 588 -----GlyTyrPheAlaGluValProGlyGlySerSerAsn 600
 DB 3058 TTTCAGAACCTGCTTTGTGGCGGAGGCTGGAAGCATCCCGCGGGGCTCTCATC 3117
 QY 601 TyrProValValTyrSerLysAspValLeuThrLeuSerGluLysPyrThrLeuTyr 620

```

Db      3118 TACCAAGTGTG-----GCCATCGACCTCGATGAGGC----- 3150
Oy      621 TTTATGpHeSerleAsnAsnGlnLysAspLutPrpLysIleGlySerValSerVal 640
Db      3151 -----CTGAAC-----GCCCTGGGTCTCTAC 3171
Oy      641 LysThrProThrGlnLutYrThrHisProLeuPheGluValGlnHisAsnGlnThrSerThr 660
Db      3172 CGCATGCCGGGGGCGATAGCCCCCGCATGAGCTTCTTCATCAACAGCAGCAGCGCGGTGG 661
Oy      661 TTTThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLys----- 677
Db      3232 CTCACCAACCAACCGAGCTGAGCCGCGAGCGCGATCGCGAGCTCGCGGTGGTGGCC 3291
Oy      678 ---AsnLeuGlyLeuProPheAsnGlnGluLeuValValPheArgGlnThrGlnSer 696
Db      3292 AGTATGACGAGCAGCGCC-----ACCAAGAC 3318
Oy      697 SerSerGlySerLeuThrAlaAlaGlnLutThrValHisIle-----LysGln 712
Db      3319 TCCACGACGACGCTC-----ACCATCATGTGCTGGATGTGAACGAGAG 727
Oy      713 GlyLutThrPhe-----ValTyrLysProValValGlnGlyProIlePro----- 727
Db      3364 ACGCCACCTCTTCCCGGCGCTGTACATGTGTGTGTGCGAGAGCGTCCACGCGAG 3423
Oy      728 -----AspGlySerTyrAlaIleThrLeuHis 736
Db      3424 TTCGGGCTGTCTGCTGACGTGACGACGACGACGAGCGGCGCTCATGCGAGCTTACG 3483-
Oy      737 AlaPheValAsnGly-----GlnGlnGlnLeuTyrLeuGlyLysArgAsnTyrThr 754
Db      3484 TACTTCATCAACAGGGGACAGTGTGAGGAACTTCAGCTGGGTACCGCATGCGCTT 3543
Oy      755 ValLysIleValAsnGly-----ThrAla-----ValGlu 764
Db      3544 GTGAGAACCGGTGGGGCTGTGACCGGAGACGACGCGCCCTACATGCTCATCTCGAG 3603
Oy      765 AlaIleGluSerSerGlnGluIleArgValPheProAsnProAlaArgAspTyrVal--- 783
Db      3604 GCCATCGACAAACGCGCTGTAGGAGACGACGACGCGGACCGCTGTCCTGCTACT 3663
Oy      784 -----GluIleSerAla 787
Db      3664 GTCTGGATGTGATGACAAACGCGCCATCTTCTGACAGAGCAGTATGAGCGCCAGCTC 3723
Oy      788 Pro---CysIleProGlnLutThrSerIleIleLeuPheAspLeuSer----- 802
Db      3724 CCTGAGACATCTCCGTGAAGCCACAGCATCTTGCAGCTGAAAGCCAGCAGCATGAG 3783
Oy      803 -----GlyLysIleValMetLysAsnSerLeuSerAlaGlnHisGly----- 816
Db      3784 GCGGAGTTTGGGCGT---GTGTGTACCGCATCTTCATGTCATGTCATGCAACACTTC 3840
Oy      817 ArgMetAspValSer-----ArgLeuPro 824
Db      3841 CGGATCCATGTCAGCAATGGCTCTCTATGCGAGGCGCGCGCGCTGACCGGAGCG 3900
Oy      825 AsnGlyAlaTyrIleLeuLysValAspGlyTyr 835
Db      3901 AACTCATCCACGCTGCTGATAGTGAAGCGCTTAC 3933

```

RESULT 11

```

US-10-090-624-5
: Sequence 5, Application US/10090624
: Patent No. US2002013235A1
: GENERAL INFORMATION:
: APPLICANT: TAKAKURA, HIKARU
: APPLICANT: MORISHITA, MIO
: APPLICANT: SHIMOTO, TOMOKO
: APPLICANT: ASADA, KIYOZO
: APPLICANT: KATO, IKUNOSHIN

```

```

: TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
: FILE REFERENCE: TAKAKURA-6
: CURRENT APPLICATION NUMBER: US/10/090,624
: CURRENT FILING DATE: 2002-03-06
: PRIOR APPLICATION NUMBER: 09/445,472
: PRIOR FILING DATE: 1999-12-06
: PRIOR APPLICATION NUMBER: 151969/1997
: PRIOR FILING DATE: 1997-06-10
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 5
: LENGTH: 4765
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Synthetic
: US-10-090-624-5

Alignment Scores:
Pred. No.: 0.000588 Length: 4765
Score: 138.00 Matches: 163
Percent Similarity: 31.13% Conservative: 87
Best Local Similarity: 20.30% Mismatches: 264
Query Match: 3.11% Indels: 290
DB: 12 Gaps: 42

US-10-030-330-1 (1-843) x US-10-090-624-5 (1-4765)
Oy      185 ProAsnGlyAsnHisAlaTyrThrGlnCysValAlaIleThrAlaAlaGlnIleMetArg 204
Db      1174 CCTAACGAGAAATGTCAGATATTGGG----- 1200
Oy      205 TyrHisSerThrProLeuGlnGlyGlySerPheAspTyrHisAlaGlySerLeuVal 224
Db      1201 -----TGGATGTACGCGTCAAGCAACT-----CAGGTAGCTGGAAGCTTCT 1245
Oy      225 Gly-----AsnTrp-----SerGly----- 229
Db      1246 GGTATGACGACGACAAATGATGCTGGATGGCTGATGCTGATGCTGATGCTGATGCTG 1305
Oy      230 ThrPheGlyGluMetTyrAspThrPheIleAsnMetProGlnAsnProAspLeuAspLeu 249
Db      1306 GTGTCTCAAGACTGATGCTGCTG-----GATTTATGCAAGCTT 1344
Oy      250 ThrGln-----SerGlnValAspAlaTyrAlaThrLeu 260
Db      1345 ACCACAGACCCGCTGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1404
Oy      261 MetArgAspValSerAlaSer----- 267
Db      1405 AGGAGTGTATGAGCGGGGTGACATGTGAGATTTATGAAAGTATGACATACGACGAC 1464
Oy      268 -----ValSerMetSerPhe-----TyrGlnAsnGlySer 277
Db      1465 CATGTGCGACGCTTAAATACATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1524
Oy      278 GlyThrTyrSerValTyrValValGlyAla-----LeuArgAsn 290
Db      1525 GATCCAGAAAGCGTGTGCTGATGA---GCTTACCGAAAGTACGCTGTGTATGCTGAT 1583
Oy      291 AsnPheArgTyrLysArgSer----- 297
Db      1584 AGCTGCGAAGAAATGAAGTCTGCGCATTAACATGCTGGAAGTCTGCTGCTGCTGCTG 297
Oy      298 -----LeuGlnLeuHisValArg-AlaLeuTyrThrSerGlnI 310
Db      1644 GGCATTAAGCTTTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1701
Oy      310 utrpHisAspMetIleArgGlyGluLeuAlaSerGlyLysProValTyrTyr----- 327
Db      1702 -----GCACCTGGAATTCCTGATTAATGATGATTTCTA 1733
Oy      328 -----AlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCysAspGlyTyrAl 344

```



```

Db      1734 TTACTCCCGCTACACAAGCTTAGATAGCATTC--TTCTCAAGCAGAGGGCGAG 1790
Oy      344 aserispglythrphenispheasntpr-----glytrpglyglyvalseransgl 361
Db      1791 AATAGATGGTGAATAAAMACCAATGTAGTGGCTCCAGCTTACGGA-----1836
Oy      361 yphetryllysleuthrleuSerProthrserleuuglyleglygluglyllegl 381
Db      1837 -----ATTACTCATCCCTGCCGATGTGGATTGGCGA-----1869
Oy      381 yphetryllyrgingluileltherrglyllegluProalalythProalaglual 401
Db      1870 -----GGGACTCATGTCTGGAACTTGATGGCT--ACTCCACATGTGAG 1913
Oy      401 aglythrspalaleuProilleuualaleuylsaspilleglualeglutryllyser-- 420
Db      1914 CGGTGTGGTGCACCTCCATAGAGGGGCGCAAGGCCGAGGAAATACTACAAATCCAGA 1973
Oy      421 -----gluserglyleuansvalglutryserlletyrasophrgl 434
Db      1974 TTAATTAAGAAGGTTCTTGAGAGCGGTCGAACCTGGCTTGAGGAGATCATATACTGG 2033
Oy      434 yglugluglnSerasnleuaspbleuglytyrarg-----445
Db      2034 GCAGAGTACTGACTGAGCTTGCACCAAGGTGCTGTTAAGCTTACCAAGTCTGGGA 2093
Oy      446 ----leuanslysalaspilglyluallegluallysthrserSerleasnillese 464
Db      2094 AATCTTAAAGCTATTAACGCCACCACTCTCCCAATTTGTATCACTGGGACAGCAAGTC 2153
Oy      464 trlpryrglytyrglygluHisProgluserpheserleualProasnlglnleusergl 484
Db      2154 CTACAGCCGACTTGGCGAGTAC-----TT 2177
Oy      484 ngllyleasntrhrllethrleuLeutyrgargtrhglythrglu-----glutry 501
Db      2178 GGGTGTGAGCGTATTAAGAGTCTCTACCAAGAACTATATACGTGACATGTGTGAGTG 2237
Oy      501 pgluProvalargHisalaglnlyglytyrValasnerllylvalasnthrthras 521
Db      2238 GCAC--ATTAAGTACGTAGGGGACACGAGTACAGAACTTTGAACTATATCAACTGA 2294
Oy      521 pPro-----AsnasnvalvalValthrValasnasnlglnludlyly 535
Db      2295 GCCATGAGATTAAAGCTTTTGTCAGTGGAGATGTAATTTAGAGAAACAATACCGAGTTTGT 2354
Oy      2355 CCTTAGGGTG-----AAATATGATGTAGAGGGGTCTGACCCAGGTCTCA 2399
Db      555 ethrValglnpheasnSeraspSerProspgluleargthrProvalalphealale 575
Oy      2400 TGTGGAGGATATCATTTGATGATCA-----ACACGCCAGTT-----2439
Db      575 userthrValalthrAlaaspaspyalileserleuuglytrpvalmetalaaglualpr 595
Oy      2440 -----ATTGAAGACGAGATCTTG-----2457
Db      595 oglylyserSerasntyrProvalaltrpserlyaspyalileuthrleuserglugl 615
Oy      2458 -----AACACAAATGTATATCCGAGAA-----TTCACCTCGTGAAG 2495
Db      615 yaspyltyrthleu----trpyrargpheserlileasnagnllylaspglutrlpysly 634
Oy      2496 CAATTACACCTTCACCTGTAT-----2517
Db      634 slleglyserValserVallysthrProthrglutrythrHisproleuPheglualgl 654
Oy      2518 -----GATATTAAATGCTCAGAAATGCTGACTCACCACTTCTTCACTGTGCC 2564
Oy      654 yHisasnlglnthrSerthrttyrThrleuasnmetalaHisasnargValleu-----Pr 672

```

```

Db      2565 TGAGAGATGAGACGTTCTCTACGCGATGACACATACATAGTGGACTAGGCTGTACAGACC 2624
Oy      672 oasp-----Phehrleuylsasnleu-----Gl 680
Db      2625 AGATGGAATGTTTGTGTCCCATACAGCATATATCTTCCGCTGACGTCAATCC 2684
Oy      680 yleuProPheasnlglyludleuValalVal-----PheargGlnthrGlnserse 697
Db      2685 AATGCCGCAAACTGGAGCTAGTATAGACTGCATTTAACTTTGACCCCTCATATAGTGC 2744
Oy      697 r-----SerglySerleuthrPalalaagl 706
Db      2745 GGGCTCTCTGTAGAGATTACGAGTACAGATGATACCAAGGTTTGGTACATTAAACAG 2804
Oy      706 urhr-----ValHislielysglnlyglutThrPheVallytyrly 719
Db      2805 GACATACCTTGACACTAACATGAAATTCCTCAATTCATTAATTAATTAATTAATTC 2864
Oy      719 sProvalalglugly-----ProileProaspGlySerTyrArgAlathrleuHisal 737
Db      2865 CCCAATTAATGCACTCTAATCCCATTTGGCTTGGACCTTACCAATCCGAGCTTGAAG 2924
Oy      737 apheValasnlglynglnlglnleuLeutyrglytyrleuylsargasntrtyrVallysil 757
Db      2925 CGTT-----GGTATGAGAGAGTCTTCATTAAGGCG-----2955
Oy      757 eValasnlglythrAlaValgluallegluserSerlglnluleargValPheProas 777
Db      2956 -----ATTGAACTCTCTGAGGACCGCGAGAGTTGAG-----2988
Oy      777 nProalargAsptyrValgluleSerAlaProCysilePro-----glnludhrseril 796
Db      2989 -----ATTAGATAGGACCAACCAAGGTTCCTCCGAATTCAGATCTAGACTT 3032
Oy      796 eileuPheaspbleuserglyllylvallelmetlysasnserleuserAlaglylsgl 816
Db      3033 GTACCTTTATGACAGTAAGGCAATTTAGTGGCTTATGATGGAAGCAACCAAGCAGAGA 3092
Oy      816 yargMetaspValserArgleuProasnnglyalalyrilleuylsValaspilgytyrth 836
Db      3093 AGAGGTTGATGAGTATCCCTTAAGCTGAGATTATTCATATAGTACATGATGTTACAG 3152
Oy      836 rthlyls 838
Db      3153 CGTACG 3159

```

```

RESULT 12
US-10-080-505-16
; Sequence 16, Application US/10080505
; Publication No. US20030073166A1
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RET/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080,505
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 4828
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (313)..(4548)
; OTHER INFORMATION:
US-10-080-505-16
Alignment Scores:

```

Pred. No.: 0.0006
 Score: 138.00
 Percent Similarity: 31.708
 Best Local Similarity: 18.778
 Query Match: 3.118
 DB: 9
 Caps: 48

US-10-030-330-1 (1-843) x US-10-080-505-16 (1-4828)

1 MetLysLys-----SerPheLeuLeuAlaIleValMetLeuPheGlyIle 15
 313 ATGAAAAAAGCTGATTTGCTGCAATTTTAAACCGCTTCATTCATTA---GGGATA 369
 16 AlameGln-----GlnHisSerAlaProValThrLysGlnArgAlaLeuSerLeu 32
 370 GATCCGACCGCTGGCGAGTCACTACT-----TAAATTTGGGAT 408
 33 AlaArgLeuAlaLeuArgGlnValSerLeuArgMetGlyGlnThrAlaValSerAspLys 52
 409 GACTACCAATATTTATGCTGATTTTCCCGAATAAAGGAAAGTTCAAGTTCAGGGCTTAA 468
 53 HisSerIleAspTyrValTyrArgGlnLysAspAlaGlnArgGlyIleThrSerGln 72
 469 ---AAATTTGAGGTGTACAAATGAAATGAAATTTAGTTGC-----ACATCAATGACA 519
 73 GlnGlySerProAlaTyrPheTyrValAlaAsnArgGlyLysAsnGlnGlyTyrAlaLeu 92
 520 AAAGCCCAATGATGATTTTCCGCTGCTCGCA-----AATGGGGTGGCGCA 570
 93 ValAlaAlaAspAspArgIleProThrIle-----LeuAlaTyrSerProIleGly 109
 571 TTGGTGGCGGATCAGTATATTTGATGCTGGCACATATAGTACCTTACCAATGTGGAT 630
 110 ArgPheAspMetAspSerMetProAspAsnLeuArgMetLysLeuGlnIle-----126
 631 TTGTGCTGTAAGACAAATAATCCGATCAACATCGTTTAACTTAAATTTGAAACGG 690
 127 -----TyrAspGlnGlnIleGlyLeuIleLeuSer 136
 691 AATATATTATAAAGCATCAACGATCCCTTATGAGAAAGACAC-----735
 137 GlnLysAlaGlnLeuAsnGlnGlnIleLeuArgThrGlnGlyValProAlaGlnValHis 156
 736 CACAAACCCAGCTTACATTAATTTGTT-----ACGAAAGCCACCCCAATCATGATGACT 789
 157 AlaLeuMetAspAsnGlnHisPheAlaAsn-----AspProMetArgTyr 171
 790 TCTGATATGAAACGCAAAATATACATAGCAGAAATATCCCGAAGCGGTGCTATC 849
 172 AsnGlnGlyTyrPro---TyrAsnAsnLysGlnProLeuLeuProAsnGlnLysHisAla 190
 850 GGTCCGGGTGGCAGTTTGGCGAAACATCA-----AACCAAGCGGCAC-----894
 191 TyrThrGlnCysValAlaThrAlaAlaGlnIleMetArgTyrHisSerTyrProLeu 210
 894 -----894
 211 GlnGlnGlnGlySerPheAspTyr-----HisAlaGlySerLeu 223
 895 CAAGTTCGGCGCATATACCTACAGACAGCAATACACAAACCAAGCGGAGCA 954
 224 ValGlnAsnTyrPheSerGlyThrPheGlnGlyMetTyrAspTyrPheLysMetProGlnLys 243
 955 GGGGCGGCTGCTCAACTGAGCGGCGAT-----GTGGCCCAAGCGGCAAT 1002
 244 ProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThrLeuMetArgAsp 263
 1003 -----TACGGCCCAT 1014
 264 ValSerAlaSerValSerMetSerPheTyrGlnAsnGlnSerGlyThrTyrSerValTyr 283
 1015 -----CCTATTCAGGCTCAACGGCGACAGCGTTCGCTATGTT-----1056

QY 284 ValValGlnValAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGlnLeuHisValArg 303
 DB 1056 -----1056
 QY 304 AlaLeuTyrThrSerGlnGlnTyrPheHisAspMetIleArgGlyGlnLeuAlaSerGlyArg 323
 DB 1057 ---ATTATGATGCGGAAAAAACAATAAGTGTATTAAGCGGCTATTTAGCGCAAC 1113
 QY 324 ProValTyrTyrAlaGlnLysAsn-----GlnSerIleGlyHisAlaPheValCys 340
 DB 1114 CCT-----TGGCGGGGACAGAGAAATACATTCACCTGCTAGCGACGCTTTTTCAT 1167
 QY 341 AspGlyTyrAlaSerAspLysPheThrPheHisPheAsnTyrPheGlyTyrPheGlyValSerAsn 360
 DB 1168 GAAATCCTTGAAATAATTTGGCTACATGCTT-----TAAAGCCATCGGCAAT 1218
 QY 361 GlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlnGlyIle 380
 DB 1219 GGTGCATACACATTAACA-----GACAAAGCGCAGCGCAGC 1254
 QY 381 GlyPheThrIleTyrGln-----GlnIleIleThrGlyIleGlnProAla 395
 DB 1255 GGCATTTCAACAACAACAAGAGAGACCATCTGAAAGTCGCGCTTAAAGACGAC 1314
 QY 396 LysThrProAlaGlnValGlyTyrAspAlaLeuProIleLeuAlaLeuLysAspIleGln 415
 DB 1315 AATTAACCTCGCGAAGGT-----AAAGACGATGTT 1344
 QY 416 AlaGlnTyrLysSerGlnSer-----GlyLeuAsnValGlyTyrSerIleTyr 431
 DB 1345 TACCAATTCACAAAGTCCAAATATATACCTCTCGTGTAAATTAAGCTGGAACCTGTT 1404
 QY 432 AsnThrGlnGlnGlnGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGly 451
 DB 1405 TTCGAGATCAAAAAAGCGCATCTTACTTTCACCAACATCAACAGGTGGCGGC 1464
 QY 452 -----GlnValIleGlnValLysThrSerSerIleAsnIleSerTyrGly 467
 DB 1465 GGTTCATTTTGAAGGTAACTTACGATATCTTCAGAAATTAAGCAACTGCGCAAGT 1524
 QY 468 TyrGlyGlnHisProGlnSerPheSerLeuAlaProAsnGlnLeuSerGlnGlyLeuAsn 487
 DB 1525 GCTGAGTGCATGATGAGACAGCT-----1551
 QY 488 ThrIleThrLeuLeuTyrArgArgThrGlyThrGln-----Tyr 501
 DB 1552 ---ACTGTTACTTGGAAAGTAAATGTGTGAAATGATCGCCTTCTAAATATGCC 1605
 QY 502 GlnProValArgHisAlaGlnGlnGlyTyrValAsnSerIleLysValAsnThrAsp 521
 DB 1606 AAAGGCACATTCACGTTAAAGCCAAAGGGAAATTAAGCTTGATGACGCTAGGCGAT 1665
 QY 522 ProAsnAsnValValValThrValAspAsnAsnGlnGly-----Lys 535
 DB 1666 GGTAAAGTCATTTTGGACGACGACGACGATCAAGCAACAACAAGCCTTATGTA 1725
 QY 536 LeuSerIleValProAsnSerPheValAlaAspLeuAsnSerTyrGlnHisSerThrIle 555
 DB 1726 ATTGGCTGTGTAGGAGAGATGACGCTTACGTTAAACGATGACAG-----1773
 QY 556 ThrValGlnPheAsnSerAsp-----562
 DB 1774 ---CAATTAATTAATTAATTTATTTTCGGCTTCGCTGCTGCTTACATCTT 1827
 QY 563 -----SerProAspGlnIleArgThrProVal 571
 DB 1828 AATGGCATTCATTAACCTTAAACCTTAACCAAAATACGATAGAGAGACGATGTT 1887
 QY 572 AlaPheAlaLeuSerThrGlnValThrAlaAspValIleSerLeuGlyTyrPheMet 591
 DB 1888 AATCAACATGCCCAACAGATCTACAGTG-----1917
 QY 592 AlaGlnValProGlyLysSerSerAsnTyrProValValTyrSerLysAspValLeuThr 611

```

Db      1918 -----ACATTACGTGCAGCGATTACATTAAAT 1944
Oy      612 LeuSerGluGlyAspTyrThrLeuTyrArgPheSerIleAsn-----AsnGln 628
Db      1945 GACACACACTGGCGATTAAACGATTAACGATTAATGCTTTAAATGCTTGGTTGGTGAAT 2004
Oy      629 LysAspGluTyrPheLysIleGly-----SerValSerValLysThrProThrGluTyr 646
Db      2005 AAAGATGATCTAAATAAATCTGACGCTTGGAATGCTTACCAATCCGCTTAACAAGAT 2064
Oy      647 ThrHisProLeuPheGluValGlnHisAsnGlnThrSerThrTyrThrLeuAsp----- 664
Db      2065 AATCACTCTCTCTATACAGGTGGAACAATTTAAAGCAATTTACTCAAGCGGTGGC 2124
Oy      665 -----MetAlaHisAsnArgValLeuProAspPhe 674
Db      2125 ACTTAACTGTTTAAAGTGTGCGCCACACACGCAATACATCATTTA-----AAT 2175
Oy      675 ThrLeuLysAsnLeuGlyLeuProPheAsnGlyGluLeuVal----- 689
Db      2176 CCGCTTAACGAGCTTGGCGACT--AGGGCGAAGTGTATTTATGATGACATTTGGATC 2232
Oy      690 -----ValPheArgGlnThrGlnSerSerGlySerLeuTyrPheAlaGlnGluThr 707
Db      2233 AACCGTACATTTAA-----GCTGAAGAAC 2256
Oy      708 ValHisIleLysGlnGlyLeuThrPheValTyrLysProVal-----ValGluGlyPro 725
Db      2257 TTCCAATTTAAAGCGGAGTACCGGTGTTCTCCGCAATGTTCTCAATTTGAAAGCAAAAT 2316
Oy      726 ---IleProAspGlySerTyrArgAlaThrLeuHisAlaPheValAsnGlyGlnGln 744
Db      2317 TGGACAATTCAGCAATTAACCCACACGACATTTGGTGTGCGCAATCAACAAATACC 2376
Oy      745 LeuTyrLeuLysGlyLysArgAsn-----TyrThrValLysIleValAsnGlyTyrAla 762
Db      2377 ATTGCACGCGCTTCAGATTTGACAGAGATTACGACTTTGAAACATTTAATTAAACCAT 2436
Oy      763 ValGluAlaIleGlnSerSerGluGluIleArgValPheProAsnProAlaArgAspTyr 782
Db      2437 AAAAAAGTTATGATTC----- 2454
Oy      783 ValGluIleSerAlaProCysIleProGlnGluThrSerIleLeuPheAspLeuSer 802
Db      2455 -----ATACCGACAACA-----CAATTAAT 2475
Oy      803 GlyLysIleValMetLysAsnSerLeuSerAlaGly---HisGlyArgMetAspValSer 821
Db      2476 GGCTCATTAATTAATTAATTAATGCAACAGCAATATTCATGCT----- 2520
Oy      822 ArgLeuProAsnGlyAlaTyrIleLeuLysValAspGlyTyrThrLysIleAsn 840
Db      2521 -----TTAGCAAAACTTAAGTAAATGCTCACTTAATTAAT 2556

```

RESULT 13

```

US-09-848-616-169
: Sequence 169, Application US/09848616
: Publication No. US20030054010A1
: GENERAL INFORMATION:
: APPLICANT: Sebbel, Peter
: APPLICANT: Dunant, Nicolas
: APPLICANT: Bachmann, Martin
: APPLICANT: Tisot, Alain
: APPLICANT: Lechner, Franziska
: TITLE OF INVENTION: Molecular Antigen Array
: FILE REFERENCE: 1700.0180002
: CURRENT APPLICATION NUMBER: US/09/848,616
: NUMBER OF SEQ ID NOS: 186
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 169
: LENGTH: 8464

```

```

: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: PFMAICDFG
US-09-848-616-169

Alignment Scores:
Pred. No.: 0.00141
Score: 138.00
Percent Similarity: 31.36%
Best Local Similarity: 17.72%
Query Match: 3.11%
DB: 9
Matches: 8464
Conservative: 141
Mismatch: 307
Indels: 403
Gaps: 53

US-10-030-330-1 (1-843) x US-09-848-616-169 (1-8464)
Oy      56 AspTyrValTyrArgGlnGlyAspAlaGluArgGlyIleThrSerGlnGluGlySer 75
Db      4535 GATGTCACATTTAATACGGCGCAGACGAGCAAGGAGTGTCCCTGCTGACACGCGCG 4594
Oy      76 ProAlaTyrPheTyrValAlaAsnArgGlyAsnAsnGluGlyTyrAlaLeuAlaAla 95
Db      4595 CAACCTGCCAGTATGGGCTGAAATACGCTTCTGCGCGATGAATCTGCTGCGGAT 4654
Oy      96 AspAspArgIleProThrIleLeuAlaTyrSerProIleGlyArgPheAspMetAspSer 115
Db      4655 GATGCTGTGTGCA-----TTAACCA 4678
Oy      116 MetProAspAsnLeuArgMetTyrPheGlnIleTyrAspGlnGluIleGlyLeuIle 135
Db      4679 ATGGTCACAGACCTACTGCGCATCTGATGCTTGTGTCAGCAGACGAGTGAACCTGAGATC 4738
Oy      136 SerGlyLysAlaGlnLeuAsnGluGluIleLeuArgThrGluGly---ValProAlaGlu 154
Db      4739 CCTCAGCAGATTTATGAGTAT-----CGCGCGCTGTGTTATATTCTCTCGAG 4786
Oy      155 -----ValHisAlaLeuMetAspAsnGlyHisPheAlaAsnAspPromet 169
Db      4787 TTATGGGATCCCGTATTAATGCCGATTCCTCAATTAATTTACAGGAAATAGTGA 4846
Oy      170 ArgTyrAsnGlnGlyTyrProTyrPheAsnLysGluProLeuLeuProAsnGlyAsnHis 189
Db      4847 CACAAATCGATGCGG-----GTAACAGCCATTAAT 4876
Oy      190 AlaTyrThrGlyValAlaThrAlaAlaGlnIleMetArgTyrHisSerTyrPro 209
Db      4877 GCATATTAAACCTACAGACTGG-----TTAATATATGGTGGCGCT 4921
Oy      210 LeuGlnGlyGluLysSerPheAspTyrHis-----AlaGlySerLeu--- 223
Db      4922 TTACGGCAGCAATACCACTGGAGTTATACAGTAGCAGACATCATCAGGTAGCAAAAT 4981
Oy      224 -----ValGlyAsnTyr----- 227
Db      4982 AATGCGACGATATCAATACCTGCTGACGAGACATTAATACGTTACCTCCGCGCTG 5041
Oy      228 -----SerGlyThrPheGlyGluMetTyrAspTyrIleAsnMetProGlyAsn 243
Db      5042 ACCTGGGTGATGATTAATCTAGCGGCAATTTTGGATTAATCTTTCGCGGCGCA 5101
Oy      244 -----ProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyrAlaThrLeuMet 261
Db      5102 CAATTTGGCTCAGATGACAAATATGTAACCGGATGCAAGAGAGATTGGCCCGGTATC 5161
Oy      262 ArgAspValSer---AlaSerValSerMetSerPheTyrGluAsnGlySerGlyThrTyr 280
Db      5162 CACGATATGCTGCTGATCTGACAGGTACATTAACAAATAGGATGATCAATTAAT 5221
Oy      281 Ser-----ValTyrValValGlyAlaLeu 288
Db      5222 AATAGTACGCTGCCACCGGCGCTTTTACCATCAACGATATCATTCGCGAGGT----- 5275
Oy      289 ArgAsnAsnPheArgTyrLysArgSerLeuGlnLeuHisValArgAlaLeu----- 305

```

Db 5276 -----ATAATGCTGACTTGCAGTAAAGATCAAAAGAGCTGACGGCAG 5320
OY 306 -----TyrThrSer-----GlnGluTrpHisAsp 313
Db 5321 ACGCAGATTTTACCCTGATCCCTATCTGCTGAGTCCCGCTTTCAGACGTGAAGGGCAT 5377
OY 314 MetIleArgGlyLeuAlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSer 333
Db 5378 ---ACTGTTATTCATTACGACAGAGAA---TACCGTAGTGAATGCGCAGCAG 5428
OY 333 -----
Db 5429 GAAAAAAGCCGCTTTTCCAGATCACTTACACAGGCTTCGCGCTGCGTGAACAATA 5488
OY 334 IleGlyHisAlaIlePheValCysAspGlyTyrAlaSerAspGlyThrPheHisAsnTrp 353
Db 5489 TATGTGGAACGCACTGGCGGATGCTTCTGCTG-----TTTAATTTC 5533
OY 354 GlyTrpGlyValSerAsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeu 373
Db 5534 GGTATCGG-----AAAACATG 5551
OY 374 Gly---IleGlyGlyGluGlyIleGlyPheThrIleTyrGlnGluIleThrGlyLe 392
Db 5552 GGGGACACTGGCGCTCTGCTGTGATGACGAGGCTAATTCACACTT-----5602
OY 393 GluProAlaLysThrProAlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLys 412
Db 5603 -----CCCGATGACAGTCAGCATGACGACAGACATCGGCTTTCCTCTAT 5647
OY 413 AspIleGluAlaGluTyrLysSerGlyLeuAsn-----ValGlyTyr---428
Db 5648 AAC-----AAATGCGTCAATGATCAGCAGCAATTTTCAGTTAGCTGCTCCGT 5698
OY 429 -----SerIleTyrAsnThrGlyGluGlnSerAsnLeuAspLeuGlyTyr 444
Db 5699 TATTCGACACAGCGATTTTAAATTTTCGCGATACAAACATACAGTGAATGAATGGCTAC 5758
OY 445 ArgLeuAsnLysAlaAspGlyGluValIleGlyValLys-----457
Db 5759 AACATTGAAACACAGCAGCA---GTTATTCAGGTTAAAGCGAAATTCACGCACTATTAC 5815
OY 457 -----
Db 5816 AACCTGCTTATTAACACGCGGAATTAACAACCTCACCGTTACTCAGCACTGGCGGC 5875
OY 458 ThrSerSerIleAsnIleSer-----TyrGlyTyrGlyGlnHisPro 472
Db 5876 ACATCAACACTGATTTGAGTGTAGCATCAAACTTATTTGGGAACGAGTAATGTGAT 5935
OY 473 GluSerPheSerLeuAlaProAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuLeu 492
Db 5936 GAGCAATTCAGCGCTGATTAATAATGCTGCTCGAAGATATCAACTGG---ACGCTCAGC 5992
OY 493 Tyr-----
Db 5993 TATACCTGACGAAAAACGCGCTGCAAAAAGAGCGGATGATGTTAGCGCTTAACGTC 6052
OY 494 -----ArgArgThrGlyThrGlnGlnIleTrpGluProValArg 505
Db 6053 AATATTCCTTTCAAGCCACTGCTGCTGCTGACAGTAATAATCTCAGTGG-----GCA 6103
OY 506 HisAlaGlnGlyTyrValAsnSerIleLysValAsn-----518
Db 6104 CATGCCAGTCCAGCTACAGCATCTACAGATCTCAACGGTCCGATGACCAATCTGGCT 6163
OY 519 -----ThrThrAspProAsnAsnValValValThrVal-----529
Db 6164 GGTGTATACGTAAGTGTCTGGAAGACAAACATCAGCTATAGGCTCAACCGCGCTAT 6223
OY 530 -----AspAsnAsnGlnGlyLysLeuSerIleValProAsnSerPheValAla 545
Db 545 -----
Db 6224 GCCGGGAGGCGATGGAATACGCA-----AGTACAGCTACGCC 6265
OY 546 AspleuAsn-----SerTyrGlnIleSerThrIle 555
Db 6266 ACGTGAAATTAATGCGGTGTTACGGCAATGCCAATATGCGTTACAGCATAC-----6319
OY 556 ThrValGlnPheAsnSerAspSerProAspGlnIleArgThrProValAlaPheAlaLeu 575
Db 6320 -----GATGATTAAG---CACTCTATTACGAGACTC 6349
OY 576 SerThrGlyAlaThrAlaAsp---AspValIleSerLeuGlyTrp-----589
Db 6350 AGCGGTGGGTACTGCTGCTATCCCAATGCGCTAAGCGTGGGAGCCGTTAAACGATAC 6409
OY 590 ---ValMetAlaGluValProGlyLysSerSerAsnTyrProValValTrpSerLysAsp 608
Db 6410 GTGGTCTCTGTTAAAGCGCTGCG-----GCAAAAGAT 6442
OY 609 ValLeuThrLeuSerGluGlyAspTyrThrLeuTrpTyrArgPheSerIleAsnAsnGln 628
Db 6443 GCA-----AAAGTCGAAACCAAG 6460
OY 629 -----LysAspGluTrpLysLysIleGlySerValSerValLysThrProThrGlu 645
Db 6461 ACGGGGTGCGTACCGACTGCGT-----GTTATGCCGTGCTTATGCCACTGAA 6514
OY 646 TyrThrHisProLeuPheGluValGlnHisAsnGlnThrSerThrTyrThrLeuAspMet 665
Db 6515 TATCG-----GAAATAGAGTGGCGGTGATACCAATACCTG 6553
OY 666 AlaHisAsn-----668
Db 6554 GCTGATTAACGCTATTAAGACGGGTGCTAACGCTTCCACCTGCGGCGATC 6613
OY 669 -----ArgValLeuProAspPheThrLeuLys 677
Db 6614 GTGCGACAGATTTAAACGCGCGCTTGGATTAACCTGCTATACCGCTACCCACCAAT 6673
OY 678 AsnLeuGlyLeuProPheAsnGlnGlyLeuValValValPheValGlnThrGlnSerSer 697
Db 6674 AATTAACCCGCTGCGCTTGG-----GGATGTGACATCAGAGATGACAGAT 6724
OY 698 SerGlySerLeu-----701
Db 6725 ACGGCTATGTTCCGATTAATGCTAGCTTAACTCAGCGGAATGCTTAAAGCGGAAA 6784
OY 702 -----TrpAlaAlaGlnGluThrValHisIle 710
Db 6785 GTTCAGGTGAATGGGAGAGAGAAATGCTGCTGCGCAATTAATCAACTGCCA 6844
OY 711 LysGlnGlyGluThrPheValTyrLysProValVal-----GluGlyPro 725
Db 6845 CCAAGAGTCCAGCAGCA---GTTATTAACCCAGCATCAGTGAATGTCGTTAAGGGGCGT 6903
OY 726 IleProAspGlySerTyrArgAlaThrLeuHisAlaPheValAsnGlyGlnGlnLeu 745
Db 6904 GATGAGAAACAACCTTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6963
OY 746 TyrLeuLysGlyLysArgAsnTyrThrValLysIleVal-----AsnGlyThrAlaVal 763
Db 6964 TTGGCTGCGGATGACAGCATTAATCTCCGCGCTATGTCAGGATTAACGGCTGATAGT 7023
OY 763 LgluAlaIleGluSerSerGluGlnIleArgValPheProAsnProAlaArgAspTyrVal 783
Db 7024 GGCGGTGATCAACCAATTTACGTGATGATGATGAAACCGGCGAGCAATTTAA 7083
OY 783 LgluIleSerAla-----ProCysIlePro-----Gly 792
Db 7084 CAACATTCGCGGACGACTCTGTTGTCATTTTCGATTTTGTGTCACCGCTGTGTA 7143
OY 792 nGluThrSerIleIleLeuPheAspLeuSerGly-----Lys 804
Db 7144 TCCGTTTCTGCGTAAGGTTGGTTACTGGCTTGCAGATACCAATGCAACCT 7203

QY 804 sileValMetLysAsnSerLeuSerAlaGlyHisGly 816
DB 7204 GCTTCACCTTGAATAATACGGTGTGACGGCTTCGGGA 7240
RESULT 14
US-09-848-616-168
Sequence 168, Application US/09848616
Publication No. US20030054010A1
GENERAL INFORMATION:
APPLICANT: Sebbel, Peter
APPLICANT: Duman, Nicolas
APPLICANT: Bachmann, Martin
APPLICANT: Tisot, Alain
APPLICANT: Lechner, Franziska
TITLE OF INVENTION: Molecular Antigen Array
FILE REFERENCE: 1700.018002
CURRENT APPLICATION NUMBER: US/09/848,616
NUMBER OF SEQ ID NOS: 186
SOFTWARE: PatentIn Ver. 2.1
EQ ID NO 168
LENGTH: 9299
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: PRIMAIDCFGH
US-09-848-616-168
Alignment Scores:
Pred. No.: 0.00163 Length: 9299
Score: 138.00 Matches: 183
Percent Similarity: 31.368 Conservative: 141
Best Local Similarity: 17.728 Mismatches: 307
Query Match: 3.118 Indels: 403
DB: 9 Gaps: 53
US-10-030-330-1 (1-843) x US-09-848-616-168 (1-9299)
QY 56 AspyrValTYrArgGlnGlyAspAlaGluArgGlyLeuThrSerGlnGluGlySer 75
DB 4535 GATGCACATTTAATACGGCGCACAGTGAACAAGGATTTGTCCTGACACGCGG 4594
QY 76 ProAlaTyRheTyValAlaAsnArgGlyAsnAsnGlnGlyTyRAlaLeuValAla 95
DB 4555 CAACGCCACGATGAGGGGTGATACGGCTTGTGCGCGGATGATGATGCGGCGAT 4654
QY 96 AspaParGileProThrIleLeuAlaTyRSerProIleGlyArgPheAspMetAspSer 115
DB 4655 GATGCCGTGTGCCA-----TTAACCCACA 4678
QY 116 MetProAspAsnLeuArgMetTrpLeuGlnIleTyRAspGlnGluIleGlyLeuIleu 135
DB 4679 ATGTTCACGAGGACGACTGCGCATGTGATGTGTGACACGAGCACTGAACCTGACGATC 4738
QY 136 SerGlyLysAlaGlnLeuAsnGlnGluIleLeuArgThrGlnGly---ValProAlaGlu 154
DB 4739 CCTCAGGCAATTTATAGTAAT-----CGGGCGCGGTGATATTTCTCTCTGAG 4786
QY 155 -----ValHisAlaLeuMetAspAsnGlyHisPheAlaAsnAspPromet 169
DB 4787 TTATGGATCCCGGATTAATATCCGATTCGCAATTATTAATTTGACGGGAATAGTGA 4846
QY 170 ArgTrpAsnGlnGlyTyRProTrpAsnAsnLysGluProLeuLeuProAsnGlnLysHis 189
DB 4847 CAGAAATCGGATTTGG-----GTTAAACGACCATAT 4876
QY 190 AlaTyRThrGlnGlyValAlaThrAlaAlaGlnIleMetArgTyRHisSerTrpPro 209
DB 4877 GCATTTTAAACCTACAGAGTGG-----TTAATAATTTGGTCCGCGCT 4921
QY 210 LeuGlnGlyGlnGlySerPheAspTyRHis-----AlaGlySerLeu--- 223
DB 4922 TTACGGCAATATCCACCTGGAGTTATTAACGTACGACGATCATCATCAAAAT 4981
QY 224 -----ValGlyAsnTrp----- 227
DB 4982 AAATGACAGATATCAATACCTGCTTGACGAGACATTAATACCGTTACGTCGCGGCTG 5041
QY 228 -----SerGlyThrPheGlnGlyGluMetTyRAspTrpIleAsnMetProGlyAsn 243
DB 5042 ACAGCTGGGATGATGATATACGAGGATATTTGATGATGATTTTAACTTTCCGCGGCA 5101
QY 244 -----ProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyRAlaThrLeuMet 261
DB 5102 CAATTGGCTCAGATATACATATATGTTACCGGATACGATACGAAGAGATTTGCCCGGATC 5161
QY 262 ArgAspValSer---AlaSerValSerMetSerPheTyRGlusAsnGlySerGlyThrTyR 280
DB 5162 CAGCGATATTCGTCGTGCTACTGCACAGCTACTATTAAACAAATGGGTGACATTTAT 5221
QY 281 Ser-----ValTyRValAlaGlyAlaLeu 288
DB 5222 AATAGTACGTCGCCACCGGCGCTTTTACCATGACGATATCTATGCCGAGGT----- 5275
QY 289 ArgAsnAsnPheArgTyRArgSerLeuGlnLeuHisValAlaArgAlaLeu----- 305
DB 5276 -----AATAGTGTGACTGTACAGTAAACATCAAGAGCTGACGCGCAGC 5320
QY 306 -----TyRThrSer-----GlnGluTrpHisAsp 313
DB 5321 ACAGATATTTTACCGTACCCCTATGCTGAGTCCGCTTTTGACAGGTGAAGGAGCAT--- 5377
QY 314 MetIleArgGlyLysLeuAlaSerGlyArgProValTyRAlaGlyAsnAsnGlnSer 333
DB 5378 ---ACTGTTATTCATTAACGGCAGAGAA-----TACCGTATGGAATATGCGCAGCAG 5428
QY 333 ----- 333
DB 5429 GAAAAAACCCGCTTTTCCAGACTACATTAATCCACGCGCTTCCGCGTGGACATA 5488
QY 334 IleGlyHisAlaPheValLysAspGlyTyRAlaSerAspGlyThrPheHisPheAsnTrp 353
DB 5489 TATGGTGAACGCACTGGCGGATCGTTATCGTGT-----TTTAATTC 5533
QY 354 GlyTrpGlyGlyValSerAsnGlnPheTyRLeuThrLeuLeuSerProThrSerLeu 373
DB 5534 GGTATCGG-----AAAAACATG 5551
QY 374 Gly---IleGlyGlyGlnGlyIleGlyPheThrIleGlyGlnGluIleThrGlyIle 392
DB 5552 GGGGACGTCGGCGCTGCTGTGTGATATGACGCGGCTAATTCACACTT----- 5602
QY 393 GluProAlaLysThrProAlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLeuLys 412
DB 5603 -----CCGATGACAGTCAAGATGACGAGCAATCGGCTTTTCTCTAT 5647
QY 413 AspIleGluAlaGluTyRLeuSerGlnSerGlyLeuAsn-----ValGlyTyR--- 428
DB 5648 AAC-----AAATCGCTCAATGATCAAGGACGCAATATTAGTAGTGGTTACCGT 5698
QY 429 -----SerIleTyRAsnThrGlnGlyGlnGluGlnSerAsnLeuAspLeuGlyTyR 444
DB 5699 TATTCGACACGCGGATATTTATTTCCGATACAAACATACAGTGAATGATGGCTAC 5758
QY 445 ArgLeuAsnLysAlaAspGlyGluValIleGluValLys----- 457
DB 5759 AACATTGAACACAGACGACGA---GTTATTCAGATTAAAGCCGAATTCACCGCATATTAC 5815
QY 457 ----- 457
DB 5816 AACCTCGCTTATTAACAACGCGGAAATTACAACCTGACTACGACCACTCGGCGC 5875
QY 458 ThrSerSerIleAsnIleSer-----TrpTyRGllyTyRGllyGlnHisPro 472
DB 5876 ACATCAACACTGATTTTGAAGTGTACCCATCAACATTATTGGGGAACGACATATGTGAT 5935

OY 626 AaAaNgLnLyAspGluTrpLysLysIleGlySerValSerValLysThrProThglu 645
Db 277 -----CAGATGAAACAAACAATGCGCGGAGCGGTGATATGACGATTAAGTCAAG 227
OY 646 TTTTThrHISProLeuPheGluValGlyHISAsnGlnThrSerThrThrLeuAspMet 665
Db 226 TTC-----AACACAGCGGCAACTCATCTTGTATTAAC 194
OY 666 AAlHISAsnArgValLeuProAspPheThrLeuLysAsnLeuGlyLeuProPheAsnGly 685
Db 193 AGCAAAAGCGGATCTCTTCCAGGCGACGAATGCTAACGCCAGCTTATATAACACGGT 134
OY 686 GluLeuValValPheArgGlnThrGlnSerSerGlySerLeuTrpAlaAlaGln 705
Db 133 GAAATGACGCTG-----GATCGCGCATTTAGTTCGGGCG--- 101
OY 706 GluThrValHISLysLysGlnGlyLysThr-----PheValTyrLysProValValGlyLys 724
Db 100 ---GCTATTCACTAACGACGATGCGCGACGAGTTTATTAAACAGGCGGTGGATGCG 44
OY 725 ProIle 726
Db 43 AAGATC 38

RESULT 16

US-10-160-758-1
Sequence 1, Application US/10160758
Publication No. US20030036076A1
GENERAL INFORMATION:
APPLICANT: EXELIXIS, INC.
TITLE OF INVENTION: CADS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE
CURRENT REFERENCE: EX02-089C
CURRENT FILING DATE: 2002-06-03
PRIOR APPLICATION NUMBER: US/10/160,758
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/328,605
PRIOR FILING DATE: 2001-10-10
PRIOR APPLICATION NUMBER: US 60/357,253
PRIOR FILING DATE: 2002-02-15
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 10065
TYPE: DNA
ORGANISM: Homo sapiens
10-160-758-1

Alignment Scores:

Seq. No.: 0.00265 Length: 10065
Percent Similarity: 136.50 Matches: 199
Best Local Similarity: 32.80% Conservative: 126
Query Match: 20.08% Mismatches: 337
DB: 9 Indels: 329
Gaps: 52

US-10-030-330-1 (1-843) x US-10-160-758-1 (1-10065)

OY 21 SerLeuProValThrLysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgGlnVal 40
Db 1030 AATCCCGGAGAGTTCAACAGCTCCGAGTACAGCGTGGCCATCAGTACGTGCACAGGTC 1089
OY 41 SerLeuArgMetGlyGln---ThrAlaValSerAspLysIleSerIleAspTyrValLys 59
Db 1090 GGGTTTGCCCTTCCACTTTCATCCAGGTGGTGACAG----- 1128
OY 60 ArgGlnGlyAspAlaGluArgGlyLysThrSerGlnGlu-----GlySer 75
Db 1129 -----GATGAGAAATTGGGCGCTGAACAGCATGTTGAGGTACTGTGGGGAC 1179
OY 76 ProAlaTyrPheTyrValAlaAsnArgGlyLysAsnGlnGlyTyrTrpAlaLeuValAlaAla 95

Db 1180 AACTCCGACACCTTCATCATCTCCGACCTCCGAGGGAAGCGGACATTCGTAT 1239
OY 96 AspAspArgTyrIleProThrIleLeuAlaTyrSerProIleGlyArgPheAspMetLys 114
Db 1240 CGGGTGGCCATCCCA-----CTGACTACGAGACCGGTGACCGTACGATTTGATCTC 1293
OY 115 -----SerMetProAspAsnLeu-----ArgMetTrpLeuGlnIleTyr 127
Db 1294 TTGGCCAAATGAGAGTGTGCTGACATGAGGCTGATGCCAAGTGAGATCATCTCATC 1353
OY 128 AspGlnGlnIleGlyLeuIleLeuSerGlyLysAlaGlnLeuAsnGlnGlnIleLeuArg 147
Db 1354 AATGAAATGACAAACCGGCCATCTTCAGCCGACGACGCTGCAACATCATCAGCTGTAG 1413
OY 148 ThrGlnGlyValProAlaGluValAlaHISAlaLeuMet-----AspAsnGly 162
Db 1414 AACGTACCCGTGGGACCTGTGCTGACAGCTCTGCAACTGCAATGATGACGAGCACC 1473
OY 163 -----HisPheAlaAsnAspPrometaArgTrpAsnGlnGlyTyrPro 176
Db 1474 TTGGGGAAGTCAGCTACTTCTTCAGTATGACCCCTGACAGTTCCTGCTG----- 1524
OY 177 TrpAsnAsnLysGluProLeuLeuProAsnGlyAsnHISAlaTyrThrGlyCysValAla 196
Db 1525 -----GACAAGGAC-----ACGGGACTCATATG 1548
OY 197 ThrAlaAla-----AlaGlnIlemetaArgTyrHISerTrpProLeuGlnGly 212
Db 1549 CTGATTCGACAGCGCGGACCTGTAGCTATCCACCGCTTCACCTGCATCATTCCTCCG 1608
OY 213 GlnGlySerPheAspTyrHISAlaGlySerLeuValGlyAspTrpSerGlyThrPheGly 232
Db 1609 GACGGGGCGCGGACGAGACCAACAGCGCGGTGACAGATCAAT----- 1650
OY 233 GluMetTyrAspTrpIleAsnMetProGlyAsnProAspLeuAspAsnLeuThrGlnSer 252
Db 1651 ---GTGTGGAT---GTCAAC-----GACAAGTCCCACTTC 1683
OY 253 GlnValAspAlaTyrAlaThrLeuMetArgAspValSerAlaSerValSerMet----- 270
Db 1684 CAAAGAGATGCTTACGTGGTCTGCTGCGGGAACAAGCTTGTGTCACACAGCTGCTG 1743
OY 271 -----SerPheTyrLysGlnGlySerGlyThrTyrSerValTyr 283
Db 1744 CGGTCGCGGGAACGATGAGATCCCTCCCAACACCAACGATCAGTACAGCATTTGTC 1803
OY 284 ValValGlyAlaLeuAlaGlnAsnPheArgTyrLysArgSerLeuGlnLeuHISValArg 303
Db 1804 AGTGCATCTGCTTTGGCAGCTACTTCGACATC----- 1836
OY 304 AlaLeuTyrThrSerGlnGlnTyrTrpHISAspMetIleArgGlyGluLeuAlaSerGlyArg 323
Db 1837 AGCCTGACGAGGCTAT-----GGAGTATCAGCTCAGTGC 1875
OY 324 ProValTyrTrpAlaGlyLysAsnGlnSerIleGlyHISAlaPheValCysAspLysTyr 343
Db 1876 CCCCTGATTTATGA----- 1890
OY 344 AlaSerAspLysThrPheHISPheAsnTrpGlyTyrGlyValSerAsnGlyPheTyr 363
Db 1891 -----CAGAAATCCAAATGGGCTGAT 1911
OY 364 LysLeuThrLeuLeuSerProThrSerLeuGlnGlyGlyGlnGlyIleGlyPheThr 383
Db 1912 TATCTGACGGTATG-----GCCATGATGCTGCAACCCCTCTCAACAGCACC 1962
OY 384 IleTyrGlnGlnIleIleThrGlyLysLeuProAlaLysThrProAla----- 399
Db 1963 GTCCCTGCACACATGAGAGTGTGATGAGAAACAACCTCCACCTTCACAGAGGCC 2022
OY 400 -----GluAlaGlyThrAspAlaLeuProIleLeu 409
Db 2023 GCTACTTGTCTCTGCTGTGAGAAACATGATGACAGAGCCACGGTGTCTTCGAT 2082


```

OY 21 SerAlaProValThrIysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgInVal 40
Db : : : : :
OY 1420 AATGCCCGGAGTCAACAGCTCCGAGTACAGCGTGGCCATCATAGCTGAGCAGCAGCTC 1479
OY 41 SerLeuArgMetCysGln---ThrAlaValSerAspLysIleSerIleAspTyrValTyr 59
Db : : : : :
OY 1480 GCGTTTGCCCTTCCACTCTTCATCCAGGTGGTGACAGC----- 1518
OY 60 ArgGlnGlyAspAlaGluArgGlyIleThrSerGlnGlu-----GlySer 75
Db : : : : :
OY 1519 ---GATAGAAATTTGGGCTGGAACAGCATTTTGAGGTGACTTGGTGGGAAC 1569
OY 76 ProAlaTyrPheTyrValAlaAsnArgLysAsnAsnGlnGlyTyrAlaLeuValAlaAla 95
Db : : : : :
OY 1570 AACTCCACCACTTCATCATCTCCCGACCTCCGTCAGAGGGAGGAGGAGCATTCGAT 1629
OY 96 AspAspArgIleProThrIleLeuAlaTyrSerProIleGlyThrPheAspMetAsp--- 114
Db : : : : :
OY 1630 CCGGTGGCCATCCCA---CTGGACTACGAGACCGTGGACCGCTACGACTTTGATCTC 1683
OY 115 ---SerMetProAspAsnLeu-----ArgMetProLeuGlnIleTyr 127
Db : : : : :
OY 1684 TTTCGCCAATGAGAGTGTGCTGACATGTGGCTATGCCAAGTGAAGTCACTTCATC 1743
OY 128 AspGlnGluIleGlyLeuIleLeuSerGlyLysAlaGlnLeuAsnGlnGluIleLeuArg 147
Db : : : : :
OY 1744 AATGAATAATGACAAACGCGCCACTCTTCAGCCAGCCACTGTACAAATCAGCCTGTACGAG 1803
OY 148 ThrGlnGlyValProAlaGluValHisAlaLeuMet-----AspAsnGly----- 162
Db : : : : :
OY 1804 AACGTACCGTGGGAGCTCTGTGTGACAGCTCTGGCAACTGACATGATGAGCAGCACC 1863
OY 163 ---HisPheAlaAsnAspProMetArgTrpAsnGlnGlyTyrPro 176
Db : : : : :
OY 1864 TTTCGGGAAATCAGTCACTTCTTCAGTGAATGACCTTCAGCAGCTTCCTGCTG--- 1914
OY 177 TrpAsnAsnLysGluProLeuLeuProAsnGlyAsnHisAlaTyrThrGlyCysValAla 196
Db : : : : :
OY 1915 ---GACAAAGGAC-----ACGGACTCATCATG 1938
OY 197 ThrAlaAla-----AlaGlnIleMetArgTyrHisSerTrpProLeuGlnGly 212
Db : : : : :
OY 1939 CTGATTCGCGAGCTGACTATGAGCTCATCCAGCGCTTCAGCTGACGATCATTCGCCG 1998
OY 213 GlnGlySerPheAspTyrHisAlaGlySerLeuValGlyAsnTrpSerGlyThrPheGly 232
Db : : : : :
OY 1999 GACGGGGCGCGGAGAGACACACGCGGCTCAGATCAAT----- 2400
OY 233 GluMetTyrAspTrpIleAsnMetProGlyAsnProAspLeuAspAsnLeuThrGlnSer 252
Db : : : : :
OY 2041 ---GTGTGGAT---GTCAAC-----GACAAAGGCGCCACTTC 2073
OY 253 GlnValAspAlaLeuTyrAlaThrLeuMetArgAspValSerAlaSerMet----- 270
Db : : : : :
OY 2074 CAGAGGATGCTTACCTGCTGCTGCGGAGAACGAGCGCTTCACACAGCTGCTG 2133
OY 271 ---SerPheTyrGluAsnGlySerGlyThrTyrSerValTyr 283
Db : : : : :
OY 2134 CCGCTCCGCGGACAGATGAAGACTCCCTCCCAACACACAGATCAGCTACACATTC 2193
OY 284 ValValGlyAlaLeuArgAsnAsnPheArgTyrLysArgSerLeuGlnHisValArg 303
Db : : : : :
OY 2194 AGTGCATCTGCTTGGACGCTACTTCGACATC----- 2226
OY 304 AlaLeuTyrThrSerGlnGluTrpHisAspMetIleArgGlyGluLeuAlaSerGlyArg 323
Db : : : : :
OY 2227 AGCTGTACGAGGCTAT-----GGAGTGTACAGCCGCTACCTCGC 2265
OY 324 ProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCysAspGlyTyr 343
Db : : : : :
OY 2266 CCCCTGATATTAGAA----- 2280

```

```

OY 344 AlaSerAspGlyThrPheHisPheAsnTrpGlyTyrGlyValSerAsnGlyPheTyr 363
Db : : : : :
OY 2281 ---CAGATATCAATGGCGTGAAT 2301
OY 364 LysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyValLeuGlyIleGlyPheThr 383
Db : : : : :
OY 2302 TATCTACGAGTCACTG-----GCCATGATGCTGGACACCCCTCTCAACAGCACC 2352
OY 384 IleTyrGlnGluIleLeuThrGlyIleGluProAlaLeuTyrProAla----- 399
Db : : : : :
OY 2353 GTCCCTGACCATTCAGAGGTGTGATGAAATGACAAACCTCCACCTTCAGCAAGCCC 2412
OY 400 ---GluAlaGlyThrAspAlaLeuProIleLeu 409
Db : : : : :
OY 2413 GCCTACTGCTCCCTGGTGAGAACATCATGTGAGAGCAGCAGCGGTGCTTCCTGAAT 2472
OY 410 AlaLeuLys---AspIleGluAlaGluTyrLysSerGlnSerGlyLeuAsnValGlyTyr 428
Db : : : : :
OY 2473 GCCACAGACCTGACCGCTCCCGGAGTACGGCCAGAGTCC-----ATCATCTAC 2523
OY 429 SerIleTyrAsnThrGlyGluGlnGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLys 448
Db : : : : :
OY 2524 TCCTTG-----CAAGCTTCACCCAG-----TTTCGATCATATGCC 2559
OY 449 AlaAspGlyGluValIle-----GluValLysThrSerIle 461
Db : : : : :
OY 2560 CGCTCAGGGGAAATCACCACACAGCTCTGTGACCGAGAGCAGCAAGCTGTGATCATC 2619
OY 462 AsnIleSerTrpTyr-----GlyTyrGlnGlnHisProGlnSerPheSerLeuAla 478
Db : : : : :
OY 2620 CTGATGTTGCGGCACTGACGGGGGTGGGACACACAGAAACT----- 2667
OY 479 ProAsnGlnLeuSerGlnGlyIleAsnThr-----IleThrLeuLeuTyrArgArgThr 496
Db : : : : :
OY 2668 ---GGCATCGCCACCGTAAATCATCACTCTCTGACATCATATGAC 2709
OY 497 GlyThrGlnGlnTrpGluProValArgHisAlaGlnGlyTyrValAsnSerIleLys 516
Db : : : : :
OY 2710 AACCAACCCACGTAAGAGACGACCCCTAC-----TACATCAACCTGGGTGAG 2757
OY 517 ValAsnThrThrAspProAsn---AsnValValThrValAspAsnAsnGlnGlyLys 535
Db : : : : :
OY 2758 ATGACCCCTCCAGACTGCTGATGACACCGAGGTGTGCTGTGACCAAGCTGGGGAG 2817
OY 536 ---LeuSerIleValProAsnSerPheValAlaAspLeuAsnSerTyr 550
Db : : : : :
OY 2818 AATGGACCTCGTGTACAGCATCCAGCCACCAACAATGCTACAGCTCAACAGCACC 2877
OY 551 ---GluHisSerThrIleThrValGlnPheAsnSerAspSerProAsp----- 565
Db : : : : :
OY 2878 ACGGCAAGATCCGACACACCCAGCCATGTGAGACGGGAGAACCCGACCCCATGAG 2937
OY 566 ---GluIleArgThrProValAlaPhe 573
Db : : : : :
OY 2938 GCCGACCTGATGCGCAAAATCGTCTGCTGTACTGACTGAGCGGACGCCCTGTGAAA 2997
OY 574 AlaLeuSerThrGlyAlaThrAlaAspAspValIleSerLeu----- 587
Db : : : : :
OY 2998 GCCACAGCAGTCCACAGATGTTTGAAACCTTGTGAACTGAAATGACAAATGACCCACC 3057
OY 588 ---GlyTrpValMetAlaGluValProGlyLysSerSerAsn 600
Db : : : : :
OY 3058 TTTCAGAACCTGCGCTTGTGGCGGAGTGTGGAAGGATCCCGGCGGGGTCTTCATC 3117
OY 601 TyrProValValTrpSerLysAspValLeuThrLeuSerGlnGlyAspTyrThrLeuTrp 620
Db : : : : :
OY 3118 TACCAAGTGTG-----GCCATGCACTGTGATGAGGCT----- 3150
OY 621 TyrArgPheSerIleAsnAsnGlnLysAspGluTrpLysIleGlySerValSerVal 640
Db : : : : :
OY 3151 ---CTGAAAC-----GGCCTGGTGTCTGAC 3171
OY 641 LysThrProThrGluTyrThrHisProLeuPheGluValGlyHisAsnGlnThrSerThr 660

```



```

OY 288 LeuArgAsnAsnPhenArgTyrLysArgSerLeuGlnLeuHisValAlaGlnLeuTyrThr 307
    :|||
    :|||
Db 6430 ATAAAGGA-----ACG 6419
OY 308 SerGlnGluThrHisAspMetIleArgGlyGlnLeuAlaSerGlyArgProValIlyTyr 327
    :|||
    :|||
Db 6418 AAAGAAAGGGGTGCTACAGCTTCAGCAAACTTGGCTAATGGAGTGAAGCC---TTA 6362
OY 328 AlaGly-----AsnAsnGlnSerIleGlyHisAlaPheValLysAspGly----- 342
    :|||
    :|||
Db 6361 GCTGGAACCTTAATTTAGTCAAAATCGGATATATTCAAAATTGATGCAACAAAACA 6302
OY 343 -----TyrAlaSerAspGlyThrPheHisPheAsnTrpGlyTyrGlyValSer 359
    :|||
    :|||
Db 6301 ATATATGTTATTTAATCATGCTGAGAGTTTGTAGCTAATATGACCTGCAATGACA 6242
OY 360 AsnGlyPheTyrLysLeuThrLeu-----LeuSerProThrSerLeuGlyIleGly----- 376
    :|||
    :|||
Db 6241 CCTGATTTTTCGGGAACGAATATTCGCTTAAGCGGTGATGCTATTCGAATTCGACAGCA 6182
OY 377 -----GlyGlnGlyIleGlyPhe-----ThrIleTyrGlnGlnIle 388
    :|||
    :|||
Db 6181 AGTTCAAGAAAGACAGACAGCATATGTTACAGTATTCGATTCGATGATTCGTAAGTT 6122
OY 389 IleThrGlyIleGluProAlaLysThrProAlaGlnAlaGlyThrAspAlaLeuProIle 408
    :|||
    :|||
Db 6121 ACTACCTCCATCAACAAATCAAAAGCAGGTGCTACACAGCAACATCATGCAATGCAACCTC 6062
OY 409 LeuAlaLeuLysAspIleGlnAlaGluTyrLysSerGlnSerGlyLeuAsnValGlyTyr 428
    :|||
    :|||
Db 6061 GTCGCAAGACCATCAATTTAAAGCAACAGC-AMCGAAACAGGAGGAGCAACCTCAGCTAT 6003
OY 429 SerIleTyr-----AsnThrGlnGlnGlnGlnSer 438
    :|||
    :|||
Db 6002 TCTGTCTCT-CTAATAATCATTCATCGTAAGAGAGAGAGCAACATGAAAAAGAAAT 5544
OY 439 AsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyGluValIleGluValLysThr 458
    :|||
    :|||
Db 5943 -----GATGGCAAGTT----- 5932
OY 459 SerSerIleAsnIleSerTrpTyrGlyTyrGlyGlnHisProGlnSerPheSerLeuAla 478
    :|||
    :|||
Db 5931 -----ATTAGTTGGTAGCCAGCTGAGGAGCGCAAGTTTGGCAGCTTTTCAGC 5884
OY 479 ProAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuLeuTyrArgTyrThr 498
    :|||
    :|||
Db 5883 -----ACAACCGGT 5875
OY 499 GlnGlnTrpGlnProValArgHis-AlaGln-----GlyGlyTyrVal-----As 513
    :|||
    :|||
Db 5874 GACCACTGGGAAATACACAGCTGACAGCAATTTGCAGCGGCGCAGCTTGCAGATGATA 5815
OY 513 nSerIleLysValAsnThrThrAspProAsnAsnValValIleThrValAspAsnGln 533
    :|||
    :|||
Db 5814 TGGCGTTCCA---AATACAGCTGATCCAGAC-----CCGACAGCAACCAATAGTAA 5767
OY 533 uGlyLysLeuSerIleValProAsnSerPheValAlaAspLeuAsnSerTyrGlnHis 553
    :|||
    :|||
Db 5766 CTTTGATTTACTTTTATTTCA-----AGCAATATGATTTTGG 5728
OY 553 rThrIleThrValGlnPheAsnSerAspSerProAspGlnIleArgThrProValAlaPhe 573
    :|||
    :|||
Db 5727 GAAATTAATCTAATTTCT-----GATGATTTTAACTAAACCAATCTCCTAA 5686
OY 573 eAlaLeuSerThrGlyAlaThrAlaAspValIleSerLeuGlyTyrValMetAlaGln 553
    :|||
    :|||
Db 5685 TAAACAGATGAAGGA---CGTGTCTGGAATGTGATCTGTGGT-----GTAGAGGA 5635
OY 593 uValProGlyLysSerSerAsnTrpProValIleTrpSerLysAspValLeuThrLeu 613
    :|||
    :|||
Db 5634 TATACAGGAGGAGCAAAAGAGCTGCGATGTGACTGCTCAATGCACTGATGAAGCTCGG 5575
OY 613 rGlnGlyAspTyrThrLeuTrpTyrArgPheSerIleAsnAsnGlnLysAspGluTrpLys 633

```

```

Db 5574 CGAGAAAGTTG----- 5562
OY 633 sLysIleGlySerValSerValLysThrProThrGluTyrThrHisProLeuPheGluVal 653
    :|||
    :|||
Db 5561 -----GAAGCAGATTTTAAACAGGAGCAATGTTTGTATTCATTA-----AT 5518
OY 653 lGlnHisAsnGlnThrSerThrTyr-----ThrLeuAspMetAlaHisAsnArgVal 670
    :|||
    :|||
Db 5517 TTTATATGAGAAACATAAAGCATATGATATCAAAATGCGTGAATCTGCGGACCAATGAC 5458
OY 670 lLeuPro-----AspPheThrLeuLysAsnLeuGlyLeuProPheAsnGlyGlu 687
    :|||
    :|||
Db 5457 CAAAACCGAAACTGTGTATACAAACAGCTTTTCAATGCTTAA-----GCGGAGATGC 5404
OY 687 uValValValPheArgGlnThrGlnSerSerSerGlySerLeuTrpAlaAlaGlnGlnLuh 707
    :|||
    :|||
Db 5403 AGTGTCTTTCGCAATGCTACAGCAGAAAGGCGCAAGATTTTGG----- 5358
OY 707 rValHisIleLysGlnGlyGluThrPheValTyrLysProValValGluGlyProIleP 727
    :|||
    :|||
Db 5357 -----CAATTTTCATTTTACCATATTTAATTAATGACAGC 5320
OY 727 oAspGlySerTyrArgAlaThrLeuHisAlaPheValAsnGlyGlnGlnLeuTyrLe 747
    :|||
    :|||
Db 5319 A-----GCTTCAATATCAAAAGCAGCTGCTACAGGC----- 5286
OY 747 uLysGlyLysArgAsnTyrThrValLysIleValAsnGlyThrAlaValAlaIleGln 767
    :|||
    :|||
Db 5285 -----AATATCACTTGGAACTGACGAGGAGCAGCATCAATTTAAAGCAACGC 5239
OY 767 uSerSerGlnGluIleArgValPheProAsnProAlaArgAspTyrValGlnLeuSerAl 787
    :|||
    :|||
Db 5238 AACAAAGAAAGAC-----AGACGAGACATACAGCGATTAAGTGC 5200
OY 787 aProCysIleProGlnGluThrSerIleLeu 798
    :|||
    :|||
Db 5199 TATGTGCTGCTCCCTAATATCATCATCAATCGTA 5166

RESULT 19
US-09-974-300-2632
; Sequence 2632, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2632
; LENGTH: 4557
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2632

Alignment Scores:
Pred. No.: 0.00186 Length: 4557
Score: 133.00 Matches: 193
Percent Similarity: 32.17% Conservative: 129
Best Local Similarity: 19.28% Mismatches: 351
Query Match: 3.00% Indels: 329
DB: 10 Gaps: 51
US-10-030-330-1 (1-843) x US-09-974-300-2632 (1-4557)

```

[illegible]

Dd	1045	AACGCTGCATTATTAGSACAAAGCAACGTTCCGGCGGAGACAAATCTGGAAACCGGCTTCTGCA	1104
Qy	338	PheVal-----CysAspGlyTyrVala	344
Dd	1105	GAGGCGCTGTGCACGCCCAAAATACGGGAAAATGATCGAAAATATGCTGCACAGGTTATTAA	1164
Qy	345	SerAspGly-----ThrPheHisPheAsnTrpGly-----	354
Dd	1165	AGCGATTCCACAGACGCTTCACGTCGACGATTTTATTAATCTATGAGAGAAAATAATCGAC	1224
Qy	365	-----TrpGlyGlyValSerAsnGlyPheTyrIleTyrIleuThrIleu-----LeuSerProthr	371
Dd	1225	GAATCCAAAGCGAGCATATACAGATCCGTTGGTTGGTTCCGCTGATTTGGCATCTTGGTTTCAGAT	1284
Qy	372	SerLeuGlyIleGlyGlyGlyGlyIleGlyPheThrIleTyrGlnGlyIleIleIleThrGly	391
Dd	1285	TCCTTAAAGGCG-----ATTCGGATTACTTTCATTCAG-----AAGCGC	1322
Qy	392	IleGlyProAlaLysThrProAlaGlyIleGlyThrAspAlaLeuProIleLeuAlaLeu	411
Dd	1324	AGTGAGCAGGCTGGACCCCTTACAGGAAGCAAGGAT-----TATACGCTC	1379
Qy	412	LysAspIleGlyAlaGlyTyrLysSerGlySerGlyLeuAsnValGlyTyrSerIleTyr	431
Dd	1372	TCTGAT-----AATGAAAGCGGGTTGAAATTAATTAACCAAGAT	1413
Qy	432	AsnThrGlyGlyGlnSerAsnLeuAsnPreuGlyTyrArgLeuAsnLysAlaAspGly	451
Dd	1414	GTTACCGGAGCTTACAA-----ATCACGTTATCAAACCGAGTCACACGGCGA	1466
Qy	452	GluValIleGlyValLysThrSerSerIleAsnIleSerTrpTyrGlyTyrGlyGlnHis	471
Dd	1462	GTAATCAATCATTAATCAACAAACCTTACGAACACTGGACGACCGGAAACAGAGAAACA	1521
Qy	472	ProGlySerPheSerLeuAlaProAsnGln-----LeuSerGlnGlyIleAsnThrIleThr	490
Dd	1522	AAGAAAGCTTCAGTACAGCCGCTTCAGAGAAATTTGATAAAGGCTACTCAACACTTGAT	1581
Qy	491	LeuLeuTyrArgArgThrGlyThrGlnGlnTrpGlyProValArgHisAlaGlnGlyGly	510
Dd	1582	-----TTCAGAGAGAA-----ACGGTCGATTGGACGATCCGATACACAAACAAACACATAC	1632
Qy	511	TyrValAsnSerIleLysValAsnThrThrAspProAsnAsnValValValThrValAsp	530
Dd	1633	ACGATGAACAATTTGGAAATTTGGATGACAAATTTCCGAATCGCGGCGCTCACGCTCTCGAC	1692
Qy	531	-----AsnAsn-----GlnGlyLys-----	535
Dd	1693	GGATCATTTCCGCTTCACAGATGTCAGATTAACAAACGTTGAAGAAAGCAAGACTAT	1752
Qy	536	-----LeuSerIleValProAsn-----SerPheValAlaAspIleu-----	547
Dd	1753	ACATTGATTAATAAAACCGGATCATGAAAGTTTCTCCCGAAACTCATCGGAGATTATTCG	1812
Qy	548	-----AsnSerTyrGlnHisSerThrIleThrValGlnPheAsnSerAspSerProAsp	565
Dd	1813	ACAACGACAGACCCAGCTTAATGATCAGTATACACAGATATGAATGACCGACTTTTCCAAC	1872
Dd	1873	GAA-----	1875
Qy	586	SerLeuGlyTrpValMetAlaGlyValProGlyGlySerSerAsnTyrProValValTrp	605
Dd	1875	-----	1875
Qy	606	SerLysAspValLeuThrLeuSerGlnGlyAspTyrThrLeuTrpTyrArgPheSerIle	625
Dd	1876	-----AATGTCAAAACACCGGCGAA-----TCAACGTGGACTGATCACAAGCAGC	1920
Qy	626	AsnAsnGlnLysAspGlyTrpLysLysIleGlySerValSerValLysThrProThrGlu	645


```

Db      16119 CTGAATAATTGACAAACAGGTGATATTATTGTAACAATGTACAGAGCGAGTGCCTAAT 16060
OY      778 ProAlaArgAspTyrValGluIleSerAlaProCys----- 789
Db      16059 ATTGGCTGCATATGGTCAGTTTAACTGAAGCAAGCGGTGATGTCATATTTCA 16000
OY      790 ---IleProGlnGluThrSerIleIleLeuPheAspLeuSerGlyLysIleValMetLys 808
Db      15999 TGGTATTCAGAAAATACCACTATCGGACTGTCGATGCATCAGGGAAAGTCACTTTGAAT 15940
OY      809 AsnSerLeuSerAlaGlyHisGlyArgMetAspValSerArgLeuProAsnGly----- 826
Db      15939 -----GTAAGGCACTGTC---GTAATTAAAGCCACATCTGGTGTATAAG 15898
OY      827 -----AlaTyrIleLeuLysValAspGlyTyrThrLysIleAsn 840
          ::::: ::::: ::::: ::::: :::::
          15897 CAACACAGTAAGTACACTATATAAGCACCGCTGCTATATGATATAAAGTGAT 15847

```

Job completed: June 2, 2003, 01:14:35
 Job time : 676 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - nucleic search, using frame_plus.p2n model

Run on: June 1, 2003, 22:41:13 ; Search time 100 Seconds
(without alignments)
2585.285 Million cell updates/sec

Title: US-10-030-330-1

Perfect score: 4438
Sequence: 1 MKKSFLLAIVMLFGIAMQGH.....PNGAVILKVDYTTKINIVH 843

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus.p2n.model -DEV=xlh
-O=/cgn2.1/USPTO/spool/US10030330/rnat_23052003_181645_6806/app_query.fasta.1.1031
-DB=Issued_Patents_NA -QEXT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=20
-MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030330.ecgn.1.1.40.etrnat.23052003_181645_6806 -NCPV=6 -ICPV=3
-NO_XLPRX -NO_MAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONCLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA: *
1: /cgn2.6/ptodata/1/lna/5A.COMB.seq: *
2: /cgn2.6/ptodata/1/lna/5B.COMB.seq: *
3: /cgn2.6/ptodata/1/lna/6A.COMB.seq: *
4: /cgn2.6/ptodata/1/lna/6B.COMB.seq: *
5: /cgn2.6/ptodata/1/lna/PCTUS.COMB.seq: *
6: /cgn2.6/ptodata/1/lna/Backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3624	81.7	8439	4	US-08-221-017B-473 Sequence 473, App
2	484	10.9	1197	5	PCT-US95-11723-7 Sequence 7, Appl
3	484	10.9	1197	5	PCT-US95-11723-7 Sequence 7, Appl
4	484	10.9	1197	5	PCT-US95-11723-7 Sequence 7, Appl
5	417	9.4	566	4	US-09-221-017B-423 Sequence 423, App
6	192.5	4.3	504	4	US-09-221-017B-129 Sequence 129, App
7	158	3.6	656	4	US-09-221-017B-37 Sequence 37, Appl
8	145	3.3	25165	4	US-09-453-702B-39 Sequence 39, Appl
9	144.5	3.3	4118	1	US-08-119-125A-3 Sequence 3, Appl
10	140	3.2	3106	4	US-08-840-466A-21 Sequence 21, Appl
11	140	3.2	3106	4	US-09-696-188B-21 Sequence 21, Appl
12	140	3.2	9432	1	US-08-277-231A-1 Sequence 1, Appl

13	140	3.2	9432	2	US-08-473-750-4	Sequence 4, Appl
14	140	3.2	9432	2	US-08-477-326-4	Sequence 4, Appl
15	140	3.2	30549	4	US-09-134-001C-322	Sequence 322, App
16	139.5	3.1	2848	2	US-08-805-918-1	Sequence 1, Appl
17	139.5	3.1	3695	4	US-09-453-702B-18	Sequence 18, Appl
18	139.5	3.1	4106	2	US-08-702-572-14	Sequence 14, Appl
19	139.5	3.1	4732	6	5521093-4	Patent No. 5521093
20	138	3.1	4765	1	US-08-750-532-8	Sequence 8, Appl
21	138	3.1	4765	4	US-08-894-818B-7	Sequence 7, Appl
22	138	3.1	4765	4	US-09-445-472-5	Sequence 5, Appl
23	137.5	3.1	4268	4	US-09-453-702B-93	Sequence 93, Appl
24	135	3.0	1666	4	US-09-221-017B-933	Sequence 933, App
25	133	3.0	43360	4	US-09-453-702B-206	Sequence 206, App
26	133	3.0	45325	4	US-09-453-702B-261	Sequence 261, App
27	132.5	3.0	3131	4	US-08-840-466A-20	Sequence 20, Appl
28	132.5	3.0	3131	4	US-09-696-188B-20	Sequence 20, Appl
29	132.5	3.0	4875	1	US-08-460-739-1	Sequence 1, Appl
30	128.5	2.9	3668	4	US-09-206-942-36	Sequence 36, Appl
31	128.5	2.9	3668	4	US-09-206-942-36	Sequence 33, Appl
32	128	2.9	2826	4	US-08-624-655A-1	Sequence 1, Appl
33	128	2.9	15512	2	US-08-853-659A-5	Sequence 5, Appl
34	128	2.9	15512	2	US-08-853-659A-63	Sequence 63, Appl
35	128	2.9	15512	2	US-08-853-659A-63	Sequence 63, Appl
36	128	2.9	15512	2	US-08-853-659A-66	Sequence 66, Appl
37	128	2.9	24701	2	US-08-853-659A-2	Sequence 2, Appl
38	128	2.9	24701	2	US-08-853-659A-3	Sequence 3, Appl
39	128	2.9	24701	2	US-08-853-659A-60	Sequence 60, Appl
40	128	2.9	24701	2	US-08-853-659A-61	Sequence 61, Appl
41	127.5	2.9	2037	4	US-08-913-942-14	Sequence 14, Appl
42	127.5	2.9	2079	4	US-09-268-347-25	Sequence 25, Appl
43	126.5	2.9	5629	4	US-09-453-702B-243	Sequence 243, App
44	126	2.8	10711	4	US-08-961-527-145	Sequence 145, App
45	125.5	2.8	8967	2	US-08-853-659A-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-09-221-017B-473
Sequence 473, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
Prior Application DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
Prior Application DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
Prior Application DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
Prior Application DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:
 NAME: Monroy, Gladys H
 REGISTRATION NUMBER: 32,430
 REFERENCE/DOCKET NUMBER: 27340-20021.00
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-813-5600
 TELEFAX: 650-494-0792
 TELEX: 706141
 INFORMATION FOR SEQ ID NO: 473:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8439 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: circular
 MOLECULE TYPE: DNA (genomic)
 HYPOHETICAL: NO
 ANTI-SENSE: UNKNOWN
 ORIGINAL SOURCE:
 ORGANISM: PORYPHYROMONAS GINGIVALIS
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 1...8439
 US-09-221-017B-473

Alignment Scores:
 Pred. No.: 0
 Score: 3624.00
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 81.66%
 DB: 4
 Length: 8439
 Matches: 684
 Conservative: 0
 Mismatches: 0
 Indels: 0
 Gaps: 0

US-10-030-330-1 (1-843) x US-09-221-017B-473 (1-8439)

QY 1 MetLysSerPheLeuLeuAlaIleValMetLeuPheGlyIleAlaMetClnIyHis 20
 Db 6386 ATCAAAAAAGTTTCTTTTACGCATAGTAATGCTTGGCATGCCATGAGGAGCAT 6445
 QY 21 SerAlaProValThrIyScIuArgAlaLeuSerIeAlaArgLeuAlaLeuArgIuVal 40
 Db 6446 TCGGTCGCGGTACGAAAGAGCGAGCTTGAGTCTGGCTCGGCTGGCTTTCGACAGGTA 6505
 QY 41 SerLeuArgMetGlyClnThrAlaValSerAspLysIleSerIleAspTyrValTyrArg 60
 Db 6506 TCCCTGCGAATGGGCAAAACACACATATCTGACAGATTTCCATCGATTAGCTTTATCGG 6565
 QY 61 GlnGlyAspAlaGluArgGlyIleThrSerGlnGluGluGlySerProAlaTyrPheTyr 80
 Db 6566 CAAAGGAGTGTGAGAGGGGTATCATCATCACAAGAGGAGGCTCTCTCGCATATTTTAT 6625
 Db 81 ValAlaAsnArgGlyAsnAsnGluGlyTyrAlaLeuValAlaAlaAspAspArgIlePro 100
 Db 6626 GTAGCTAATCGTGAAATATATGAGGCTATGCTCTTGTACACAGATGACAGAAATACCG 6685
 QY 101 ThrIleLeuAlaTyrSerProIleGlyArgPheAspMetAspSerMetProAspAsnLeu 120
 Db 6686 ACAATTTTATGCTTATACCCATGAGGCTTTCGACATGACAGATMTGCCGACATCTT 6745
 QY 121 ArgMetTyrLeuGlnIleTyrAspGlnGluIleGlyLeuIleLeuSerGlyLysAlaGln 140
 Db 6746 CGCATGTGGCTACAAATTTAGCATACAGAAATAGGCTGTATCTTCCGAAAAGCTCAG 6805
 QY 141 LeuAsnGluGluIleLeuAlaGlnGluGlyValProAlaGluValAlaHisAlaLeuMetAsp 160
 Db 6806 CTCAATGAAAGATATTATACGTACCGAGGCGTACCGCTGAGTACATCTCTGTATGAT 6865
 QY 161 AsnGlyHisPheAlaAsnAspPheMetArgTyrPheAsnGlnGlyTyrProTyrAsnAsnLys 180
 Db 6866 AACGGTCATTTTCCCAACGATCCCATGCAATGCAAGGTTTACCCATGGAACAATTAAG 6925
 QY 181 GluProLeuLeuProAsnGlyAsnHisAlaTyrThrGlyCysValAlaThrAlaAla 200
 Db 6926 GAACCACTGCTTCTTAATGCAATCATGCTATACGGCTGTGTGCTACTGCTGACAGCA 6985

QY 201 GlnIleMetArgTyrHisSerTyrProLeuGlnGluGlySerPheAspTyrHisAla 220
 Db 6986 CAAATCATGCGGTACCATATGCTGCGGCTTCAAGGTGAAGCTCTTTCGATTATCATGCA 7045
 QY 221 GlySerLeuValGlyAsnTyrSerGlyThrPheGlyGluMetTyrAspTyrIleAsnMet 240
 Db 7046 GGTCTATTATGTGGCAACGTGCTCCGACATTTGGTGAATGTACGATGATCAATATG 7105
 QY 241 ProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAlaAlaTyrAlaThrLeu 260
 Db 7106 CCGGAAAATCCGACCTTGATATATCTCAATCTCAAGGATGCTTACGCGACACATG 7165
 QY 261 MetArgAspValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThrTyr 280
 Db 7166 ATGCGTATGTGAGGCACTGTGTTCCATGATGATGATTTATGAAATGGAAGTGTACTGAC 7225
 QY 281 SerValTyrValValGlyAlaLeuArgAsnAsnPheArgTyrTyrAspSerLeuGlnLeu 300
 Db 7226 AGCGTTATGTATGATGAGGAGGCTTCCGAAACACTTTCGACAAAGGCTTACGACGCTA 7285
 QY 301 HisValArgAlaLeuTyrThrSerGlnGluTyrPheAspMetIleArgGlyGluLeuAla 320
 Db 7286 CATGTACGCGGCTTATATACCTCACAGAGTGCACAGATATGATCCGCGGGAACCTTGC 7345
 QY 321 SerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCys 340
 Db 7346 TCCGGAAGGCGGCTATATATGACAGGAATACCAACGATAGACATGCTTTCGTTGC 7405
 QY 341 AspGlyTyrAlaSerAspGlyThrPheHisPheAsnTyrPglyTyrGlyValSerAsn 360
 Db 7406 GATGTTATATCTTCCGATGATGATCTTCCATTTCACTGGGCTGGGAGGTGTTCCAC 7465
 QY 361 GlyPheTyrIleuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGluTyrIle 380
 Db 7466 GCGTTTACAACTAACACTCTCTCCGCACTTGGTGGATCGAGGTGAGGAGATA 7525
 QY 381 GlyPheThrIleTyrGlnGluIleIleThrGlyIleGluProAlaLysThrProAlaGlu 400
 Db 7526 GGTTTTACATTTATCAAGATATCAACCGGTATCGAACCGGTAAAGCTCCGCGTAA 7585
 QY 401 AlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGluValGluTyrLysSer 420
 Db 7586 GCCGATACAGAGCTTCCGATCTTGGCACTGAAGACATAGAGCGAGTATTAAGT 7645
 QY 421 GluSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGlyGluGlnIleAsnLeu 440
 Db 7646 GAATCGGATTAACGATAGGATATTCATATATATACAGTGAAGGCAATCAATCTT 7705
 QY 441 AspLeuGlyTyrArgLeuAsnLysAlaAspGlyGluValIleGluValLysThrSerSer 460
 Db 7706 GACCTCGATACAGATTAACAAGGCTGACGAGGAGTCAATAGAGTGAACCTTATCT 7765
 QY 461 IleAsnIleSerTyrPyrGlyTyrGlyGluHisProGluSerPheSerLeuAlaProAsn 480
 Db 7766 ATCAATATCTTGTGATCGATACGAGAGCACTCCGAAAGTTCATTTGCGACCTAAT 7825
 QY 481 GluLeuSerGlnGlyIleAsnThrIleThrLeuLeuTyrArgTyrGlyThrGlnGln 500
 Db 7826 CAGTGTCAACAAGAAATCAACACATCACCTTATCGTCGACAGGCAAGCCGAAACAG 7885
 QY 501 TyrGluProValArgHisAlaGlnGlyGlyTyrValAsnSerIleLysValAsnThrThr 520
 Db 7886 TGGAGCGGCTACGCGATGCACAGGAGGATATGTCATATAGCATTAAGTAAATACGACA 7945
 QY 521 AspProAsnAsnValValValThrValAspAsnAsnGluGlyLysLeuSerIleValPro 540
 Db 7946 GACCCGAAACAGTGTCTATGACGCTAGATATATACGAAGCAAGCTCAGATGCTCC 8005
 QY 541 AsnSerPheValAlaAspLeuAsnSerTyrGluHisSerThrIleThrValGlnPheAsn 560
 Db 8006 AACAGCTTGTGCGAGATCTGAATTTCTATATGAAATATGATACAGTATACAGTTCAT 8065

Oy	561	SerSSerSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrGlyAlaThr	580
Db	8066	AGCGACACCCCTGTATGCATCTCGACACCCGTAGCCTTTGGCTCATATCACAGGAGCTACT	81255
Oy	581	AlaAspAspValIleSerLeuGlyTTPValMetAlaGluValProGlyGlySerSerAsn	600
Db	8126	GGGAGCATGTAAATATCTTTGGCTGGGTATGGCTGAAGTTCCGGGGCGGTAGCAGCAAC	81855
Oy	601	TyrProValValITripSerLysAspValLeuThrLeuSerGluGlyAspTyrThrLeuTP	620
Db	8186	TATCCGGGGGTTTGGTCTAAAGACGTTCTCACTCTCGAAGGCGCACTATACATTTGGG	8245
Oy	621	TyTrpPheSerIleAsnAsnGlnLysAspValuTPryLysIleGlySerValSerVal	640
Db	8246	TATGATATTTTCATCATCAACACCAAAAGGATGAATGAAAAAGATCGGAAGCGGTGCAGTA	8305
Oy	641	LysThrProThrGluTyrThrHisProLeuPheGluValGlyHisAsnGlnThrSerThr	660
Db	8306	AAAACACCGACAGAGTATAGCCACCCTTATTCGAATGGGGCCATTAATCAAACTTCAC	83655
Db	8366	TATACGGTGGATATGGCAGACACACAGAGTATTCGCCGACTTACACATCAAAATCTCGGA	84255
Oy	681	LeuProPheAsn 684	
Db	8426	TTGGCTTTCAAT 8437	

```

1  RESULT 2
2  US-08-931-220-7
3  : Sequence 7, Application US/08931220
4  : Patent No. 6030835
5  :
6  : GENERAL INFORMATION:
7  : APPLICANT: Musser M.D., James M.
8  : APPLICANT: Kapur M.D., Vivek
9  : TITLE OF INVENTION: Methods and Compositions for Identifying
10 : TITLE OF INVENTION: Group A Streptococcus
11 : NUMBER OF SEQUENCES: 58
12 : CORRESPONDENCE ADDRESS:
13 : ADDRESSEE: WELLS, GOTSHAL & MANGES
14 : STREET: 2882 Sand Hill Road, Suite 280
15 : CITY: Menlo Park
16 : STATE: CA
17 : COUNTRY: USA
18 : ZIP: 94025-7022
19 :
20 : COMPUTER READABLE FORM:
21 : MEDIUM TYPE: Floppy disk
22 : COMPUTER: IBM PC compatible
23 : OPERATING SYSTEM: PC-DOS/MS-DOS
24 : SOFTWARE: Patentin Release #1.0, Version #1.30
25 : CURRENT APPLICATION DATA:
26 : APPLICATION NUMBER: US/08/931,220
27 : FILING DATE:
28 :
29 : CLASSIFICATION: 435
30 : PRIOR APPLICATION DATA:
31 : APPLICATION NUMBER: US 08/306,542
32 : FILING DATE: 14-SEP-1994
33 : APPLICATION NUMBER: US 08/160,965
34 : FILING DATE: 02-DEC-1993
35 : ATTORNEY/AGENT INFORMATION:
36 : NAME: Rae-Venter Ph.D., Barbara
37 : REGISTRATION NUMBER: 32,750
38 : REFERENCE/DOCKET NUMBER: BAYL-004/01US
39 : TELECOMMUNICATION INFORMATION:
40 : TELEPHONE: (415) 926-6200
41 : TELEFAX: (415) 854-3713
42 : INFORMATION FOR SEQ ID NO: 7:
43 : SEQUENCE CHARACTERISTICS:
44 : LENGTH: 1197 base pairs
45 : TYPE: nucleic acid
46 : STRANDEDNESS: single
47 : TOPOLOGY: linear
48 : MOLECULE TYPE: cDNA
49 :

```

[illegible]

```

Db      220 -----TCTGGCTCTAATATGATATGTTTACAAATATTTCTACTGGA 258
OY      86 AsnAsnGluGlyTyrAlaLeuValAlaAlaAspAspArgTlleProThrIleLeuAlaTyr 105
Db      259 -----GGATTTTGTTATCGTTTCAGAGATAAACGTCCTCCACAAAATTCNAGATAC 309
OY      106 SerProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMetTrpLeuGln 125
Db      310 TCTACCAGCGGATCATTTTGACCTTAACGCT---AAAGAAAACACTGCTCTTCCTCATGGAA 366
OY      126 IleTyr-----AspGlnGluIleGlyLeuIleLeuSerGlyLysAla 139
Db      367 AGTTATGTCGAAACAATCAATCAAGAAAATAATAGACACTCTATGCTGTACCGCT 426
OY      140 GlnLeuAsnGluGluIleLeuArgThrGluGlyValProAlaGluValHisAlaLeuMet 159
Db      427 GAGATTAAACAACGAGTT-----GTAAATCTTCCTT 459
OY      160 AspAsnGlnHisPheAlaAsnAspPromeTarGTrpAsnGlnGlyTyrProTrpAsn 179
Db      460 GAT-----TCAAAAGCAATTCATTACACCAAGTAACCTTCAACCTA 504
OY      180 LysGluProLeuLeuProAsn-----GlyAsnHisAla 190
Db      505 TTGACACCGCTGTTATGAAAAAAGTAAACCGAGTGAAACAATCTTTGTAGTCAACATGCA 564
OY      191 TyrThrGlyCysValAlaThrAlaAlaGlnIleMetArgTyrHisSerTrpProLeu 210
Db      565 GCTACAGAGATGTGTGCTACTGCTCAACCTGCTCAAAATTATGAAATATCATATTAACCT 621
OY      211 GlnGlyGluGlySerPheAspTyr----- 218
Db      622 ---AACCAAGGTTGAAAGACTACACTAGACACTAAGCTCAAAATTAACCATATTTCAAC 678
OY      219 HisAlaGlySerLeuValGlyAsnTrpSerGlyThrPheGlyCyluMetYrAspTrpIle 238
Db      679 CATCCCTAAGAACTTG-----TTTGGAGCTATCTCTACTAGACAAATCAACCATGGAAC 729
OY      239 Asn---MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAlaAspAlaTyr 257
Db      730 AACATCCCTACCTTATATAGCGGAAGAGAACTAATACGTTCAAAAAATG-----GCGATT 783
OY      258 AlaThrLeuMetArgAspValSerAlaSerValSerMetSerPheTyrGluAsnGlySer 277

```

Db 784 TCAGAAATGATGGCTGATGTTGATTTACGATGACATGATTTATGCTCATCTAGTGT 843
QY 278 G1YThrTyrSerValIYrValIValIAlaLeuAAsnAspPheArGlyLysArgSer 297
Db 844 TCAGCAAGTACGCTGCTGCTGCTCAAGAGCCTTGAAAGAAACTTGGCTACAAACCAATCT 903
QY 298 LeuGlnLeuH1sValArGAlaLeuTyrTrpSerGlnGluTrpHisAspMetLeArgGly 317
Db 904 GTTCACCAATATMACCGTACGCGACTTACCAACAGATGGAGACCAATATGACAA 963
QY 318 GluLeuAlaSerGlyArProValIYrTyrAlaGlyAsnAsnGlnSerIleGlyHisAla 337
Db 964 GAATATATCAAAACCAACCACTATACCAAGCTGCTGGTAAAGTGGGAGCAATGCC 1023
QY 338 PheValCyAspGlyTyrAlaSerAspGlyTrpPheHisPheAsnTrpGlyIleGly 357
Db 1024 TTGTATTCAGATGCTGCTGACGAGCTACCTTACCACTGTTAACTGGGCTGGGAGTGA 1083
QY 358 ValSerAsnGlyPheTrpTyrLysLeuThrIleuLeuSerProThrSerLeuGlyIleGly 377
Db 1084 GTCCTGACGCTTCTTCCGCTGCTGACGCACTAAACCTTCAGCTTGGTACTGTGTGC 1143
QY 378 GluGlyIleGlyPheThrIleTyrGlnGluIleThrGlyIleGlyPro 394
Db 1144 GCGCAGCGCGCTTCAACGCTTACCAAGTCTGTGTAGCACTAACT 1194

RESULT 3
PCT-US95-11723-7
Sequence 7, Application PC/TUS9511723
GENERAL INFORMATION:
APPLICANT: Musser M.D., James M.
APPLICANT: Kapur M.D., Vivek
TITLE OF INVENTION: Methods and Compositions for
NUMBER OF INVENTION: Identifying Streptococcus
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAW OFFICES OF BARBARA RAE-VENTER
STREET: P.O. Box 60039
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/11723
FILING DATE: 14-SEP-1995
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/306,542
FILING DATE: 14-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Rae-Venter Ph.D., Barbara
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: BAYL-004/0305
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 926-6205
TELEFAX: (415) 424-8760
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1197 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
STRAIN: MGAS 1719

IMMEDIATE SOURCE:
CLONE: SpeB7 (cysteine protease)
PCT-US95-11723-7

Alignment Scores:
Pred. No.: 17e-41
Score: 484.00
Percent Similarity: 49.628
Best Local Similarity: 31.498
Query Match: 10.918
Gaps: 13

US-10-030-330-1 (1-843) x PCT-US95-11723-7 (1-1197)

QY 26 LysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgGlnValSerLeuArgMetGly 45
Db 107 MAGAACCAAGATAGCGCTATACATTTATCCAAATAATCAGCA-GCTATCAAGCAGT 165
QY 46 GlnThrAlaValSerAspLysIleSerIleAspTyrValIYrArGlnGlyAspAlaGlu 65
Db 166 GCACGAGCGCAGAGAT--ATTAGCTTACAAAGTTAAGTGAAGTGAAGT-- 219
QY 66 ArgGlyIleThrSerGlnGlnGlySerProAlaTyrPheTyrValAlaAsnArgGly 85
Db 220 -----TCTGGCTCTATATGATGTATATCATATTTCTACTGGA 258
QY 86 AsnAsnGlnGlyTyrAlaLeuValAlaAlaAspAspArgIleProThrIleLeuAlaTyr 105
Db 259 -----GGATTTGTATTCGCTTTCAGAGATTAACGTTCTCCAAATAATTCAGATAC 309
QY 106 SerProIleGlyArPheAspMetAspSerMetProAspAsnLeuArgMetTrpLeuGln 125
Db 310 TCTACAGCGGATCATTTGACGCTAACGGT--AAAGAAACATTCCTTCCATGAGAA 366
QY 126 IleTyr-----AspGlnGlnIleGlyLeuIleLeuSerGlyLysAla 139
Db 367 AGTATGTCGACAAATCAAGAAACAAATAATGACATACCTATTCGCTGACCGT 426
QY 140 GluLeuAsnGlnGluIleLeuArgThrGlnGlyValProAlaGlnValHisAlaLeuMet 159
Db 427 GAGATTAAACACACAGT-----GTAAATCTCTCTT 459
QY 160 AspAsnGlyHisPheAlaAsnAspPrometArgTyrAsnGlnGlyTyrProTyrAsnAsn 179
Db 460 GAT-----TCAAAAGCATTCATTACACCAAGTACCCCTTACCACTGA 504
QY 180 LysGluProLeuLeuProAsn-----GlyAsnHisAla 190
Db 505 TTGACACCTGTATATGAAAGTAAACAGGTGAACAATCTTTGTAGTGACATCA 564
QY 191 TyrThrGlyCyValAlaThrAlaAlaGlnIleMetArgTyrHisSerTrpLeu 210
Db 565 GCTACAGATGCTGTGCTACCTCACTCAATATGAATATCAATTAATTAATCCCT-- 621
QY 211 GlnGlyGlnGlySerPheAspTyr----- 218
Db 622 ---AACCAAGGCTGAAAGCTACACTAGACACTAAGCTCAANTATACCATATTTCAC 678
QY 219 HisAlaGlySerLeuValGlyAsnTrpSerGlyThrPheGlyMetTyrAspTrpIle 238
Db 679 CATCTAGACACTG-----TTTGACGCTATCTCTACTAGACATCAACTGAGAC 729
QY 239 Asn---MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyr 257
Db 730 AACATCTCTACTACTATAGCGAAGAGATTCACGTTCAAAAATG-----GCGATT 783
QY 258 AlaThrLeuMetArgAspValSerAlaSerValSerMetSerPheTyrGlnGlySer 277
Db 784 TCAGAAATGATGGCTGATGTTGATTTACGATGACATGATATGCTCATCTAGTGT 843
QY 278 G1YThrTyrSerValIYrValIValIAlaLeuAAsnAspPheArGlyLysArgSer 297
Db 844 TCAGCAAGTACGCTGCTGCTGCTCAAGAGCCTTGAAAGAAACTTGGCTACAAACCAATCT 903

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

IMMEDIATE SOURCE:		
CLONE: speB7 (cysteine protease)		
PCT-US96-05997-3		
Alignment Scores:		
Pred. No.:	1,7e-41	Length: 1197
Score:	484.00	Matches: 125
Percent Similarity:	49.62%	Conservative: 72
Best Local Similarity:	31.49%	Mismatches: 138
Query Match:	10.91%	Indels: 63
DB:	5	Gaps: 13
US-10-030-330-1 (1-843) x PCT-US96-05997-3 (1-1197)		
QY	26	LysGIuARgAlaLeuSerIleuAlaArgLeuAlaLeuArgGluIValSerIleuArgMetGly
DB	107	AAGAAGCAAAAGATATAGCGCTATCATCTTATATCCAAATAATCAGCA-GCTATCAAAAGCAGT
QY	46	GIuThrAlaValSerAspLysIleSerIleAspTyrValTyrArgGluIValAspAlaGlu
DB	166	GCACGACGCGCAGGAAGAT--ATTAGCTTGTACAAAGATTAACTTACGTGAGAACCTT--
QY	66	ArgGIyIleThrSerGluIuGluSerProAlaTyrPheTyrValAlaAsnArgGly
DB	220	-----TCTGGCTCTAATATGATATGTTTACATATATTTCTACTGCA
QY	86	AsnAsnGluGIyTyrAlaLeuValAlaAlaAspAspArgIleProThrIleLeuAlaTyr
DB	259	-----GGATTTGTTATCGTTTCCAGGAGATAAACGTTCTCCAGAAATATGAGTAC
QY	106	SerProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMetIlePleuGln
DB	310	TCTACACAGCGATCATTTGACGCTAACGCT--AAAGAAACATTCCTCTCCATCAGAA
QY	126	IleTyr-----AspGluIuIleGlyLeuIleLeuSerGIyLysAla
DB	367	AGTTATGTGCGAACAAATCAAAAGAAAATAATTAGACACTATATGCTGGTACCCT
QY	140	GluLeuAsnGluGIuIleLeuArgThrGIuGlyValProAlaGluIValHisAlaLeuMet
DB	427	GAGATTAAACACACAGTT-----GTTAAATCTCTCCTT
QY	160	AspAsnGIyHisPheAlaAsnAspProMetIaGTrPAsnGluGIyTyrProThrAsnAsn
DB	460	GAT-----TCAAAAGCATTCATTTACACCAAGGTAACCTTACCAACTTA
QY	180	LysGIuProLeuIleProAsn-----GlyAsnHisAla
DB	505	TTGACACCGGTATGTAAGAAAGTAATAAACCAAGCTGAACAATCTTTGTATGTAACAATGCA
QY	191	TyrThrGIyCysValAlaThrAlaAlaGluIleMetArgTyrHisSerTyrProLeu
DB	565	GCTACAGATGTGTTGCTCTACTGCCAAGCTGTCAAATATATGAAATATCATATTAACCT--
QY	211	GIuGIyGluGIySerPheAspTyr-----
DB	622	---AACCAAGGTTGAAGACTACACTTAGACACTAAGCTCAAAATTAACCATATTTCAAC
QY	219	HisAlaGIySerLeuValGIyAsnTyrPserGIyThrPheGIyGluMetTyrAspTyrIle
DB	679	CATCCTAAGAACTTG-----TTTGGAGCTATCTCTACATAGAACATAACACTGGAAC
QY	239	Asn---MetProGIyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAlaTyr
DB	730	AACATCTTACTACTTATATGCGGAAGACATCTAACGTTCAAAAAATG-----GCGATT
QY	258	AlaThrIleuMetArgAspValSerAlaSerValSerMetSerPheTyrGluAsnGIySer
DB	784	TCGAATATGATGGCGATGTTGTGATTTTCAGTAGACATGATATATGTCATCTAGTGT
QY	278	GlyThrTyrSerValTyrValGIyAlaLeuArgAsnAsnPheArgTyrIysArgSer
DB	844	TCTGAGAGTACTCTTCGTGTTCAAGAGCGCTGAAAGAAATTTTGGCTTACCAACATCT

QY 298 LeuGlnLeuHisValArgAlaLeuTyrThrSerGlnGluTyrPheHisAspMetIleArgGly 317
Db 904 GTTCACCAATTAACGGTTCAGCTTACCAACAGAGTGGGAACACCAATTTGCAAA 963
QY 318 GluLeuAlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAla 337
Db 964 GAATTTCTCCAAACCAACAGATACCAAGGTTCGCGTAAAGTGGGACATGCTCC 1023
QY 338 PheValCysAspGlyTyrAlaSerAspGlyThrPheHisPheAsnTyrGlyTyrPheGly 357
Db 1024 TTGTATGATGATGCTGCTCAGCAGCTTACCAATGTTAACTGGGTTGGGCTGA 1083
QY 358 ValSerAsnGlyPheTyrLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGly 377
Db 1084 GTCTGTGAGCGCTTCTTCCTCTGACGCACTAAACCTTACGCTTGTGCTGTCG 1143
QY 378 GluGlyIleGlyPheThrIleTyrGlnGluIleIleThrGlyIleGluPro 394
Db 1144 GCGCAGCGCGCTTACAGCTTACCAAGTCTGTGAGCGCACTAAACCT 1194
US-09-221-017B-423
Sequence 423, Application US/09221017B
Patent No. 644799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: IBM Compatible
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H.
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 423:
SEQUENCE CHARACTERISTICS:
LENGTH: 566 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO

AMTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc feature
LOCATION: 1...566
US-09-221-017B-423
Alignment Scores:
Pred. No.: 6e-35
Score: 417.00
Percent Similarity: 62.508
Best Local Similarity: 48.308
Query Match: 9.408
DB: 4
Gaps: 2
US-10-030-330-1 (1-843) x US-09-221-017B-423 (1-566)
QY 276 GlySerGlyThrTyrSerValTyrValAlaGlyAlaLeuArgAsnAsnPheArgTyrIys 295
Db 1 GGCAGTGCACCTTCTCCATTTTTCAGACGCCCTCGCGAGACCTTTCATTACAG 60
QY 296 ArgSerLeuGlnLeuHisValArgAlaLeuTyrThrSerGlnGluTyrPheHisAspMetIle 315
Db 61 AATCGCTTCGTACATCCATCGCTCTCCCTCCGCTAGAGTAAGTAAGTAAGTACATGAT 120
QY 316 ArgGlyIleLeuAlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIleGly 335
Db 121 CETAAGAACTGGCAGAAACAGACAGCTGTATTAATGCGGTGAGCAGGATCGATCGGA 180
QY 336 HisAlaPheValCysAspGlyTyrAlaSerAspGlyThrPheHisPheAsnTyrGlyTyrP 355
Db 181 CATGCTTCGTTTCGATGATGATACCAACAGCAGCAAGTTCACCTTCACTTGGGATGG 240
QY 356 GlyGlyValSerAsnGlyPheTyrLeuThrLeuLeuSerProThrSerLeuGlyIle 375
Db 241 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 376 GlyGlyIleGlyIleGlyPheThrIleTyrGlnGluIleIleThrGlyIleGluProAla 395
Db 301 GCGCAGCGATGAGGCTACTCTACTGACCAAGAGTGTCTATGACATAGAGCCGCCG 360
QY 396 LysThrProAlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGlu 415
Db 361 AGCAATGAAGTC---CCCGTATGCTACCGGATCCACCACTACGCTTATGATGATGCA 417
QY 416 AlaGluTyrIysSerGlnSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGlyGlu 435
Db 418 ---CACATATGCTGTGATGAGCCTGACCTGAGTGAATAAATCAAGAACTACATCCACA 474
QY 436 GluGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnValAlaAspGly 451
Db 475 TATCAGAGGATGTGAATTTGGCTTATGCTGACGCTTCCCAATGGA 522
RESULT 6
US-09-221-017B-129/C
Sequence 129, Application US/09221017B
Patent No. 644799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: Windows

```

Db      111 TTCCCGGAGGTCATCGATATGCTTTCACAGGGGCACTGCATCGGCGCGTATACGGAC 53
Oy      119 AsnLeuàrgmetrPleugInlletYrAspGIngluIle 131
          |||||  |||||  |||||  |||||  |||||  |||||
Db      51 AATCTCAGAGGTGGCTCAAGGTTATGACGCTGAATG 13

RESULT 7
US-09-221-017B-37
: Sequence 37, Application US/09221017B
: Patent No. 6444799
: GENERAL INFORMATION:
: APPLICANT: Ross, Bruce C.
: TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
: NUMBER OF SEQUENCES: 1120
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORRISON & FOERSTER
: STREET: 755 PAGE MILL ROAD
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304-1018
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows
: SOFTWARE: FastSeq for Windows Version 2.0b
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/221,017B
: FILING DATE: 23-DEC-1998
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PP1182
: FILING DATE: 31-DEC-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PP1546
: FILING DATE: 30-JAN-1998
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PP2911
: FILING DATE: 09-APR-1998
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/AU98/01023
: FILING DATE: 10-DEC-1998
: ATTORNEY/AGENT INFORMATION:
: NAME: Monroy, Gladys H
: REGISTRATION NUMBER: 32,430
: REFERENCE/DOCKET NUMBER: 27340-20021.00
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-813-5600
: TELEFAX: 650-494-0792
: TELEX: 706141
: INFORMATION FOR SEQ ID NO: 37:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 656 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: circular
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: UNKNOWN
: ORIGINAL SOURCE:
: ORGANISM: PORYPHYROMONAS GINGIVALIS
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: 1...656
US-09-221-017B-37

Alignment Scores:
Pred. No.: 1,81e-07 Length: 656
Score: 158.00 Matches: 53
Percent Similarity: 41.95% Conservative: 46
Best Local Similarity: 22.46% Mismatches: 101
Query Match: 3.56% Indels: 36
Gaps: 7

```

US-10-030-330-1 (1-843) x US-09-221-017B-37 (1-656)

```

QY 598 SerSerAsnTyrProValValTyrSerLysAspValLeuThrLeuSerGluGlyAspTyr 617
DB 9 GAGAGCAATATCATCTGTCTGTACGGGCGATATGAAATCGCTTATGAA----- 59
QY 618 ThrLeuTyrPyrArgPheSerIleAsnAsnGlnLysAspGluTyrLysIleGlySer 637
DB 60 -----GATGCCCTATGATAGAGTTCGCTTCC 86
QY 638 ValSerValLysThrProThrGluTyrThrHisProLeuPheGluValGlyHisAsnGln 657
DB 87 ATAGAAGTACGACAGACATACGACACCATTCATCTGTTACTGAGGCGCTCCACCCACAG 146
QY 658 ThrSerThrTyrThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLys 677
DB 147 ATCGATCTTCCACAGTTCATCGGCGCAATCCGACATTCGCGCTCCACCATTCAC 206
QY 678 AsnLeu---GlyLeuProPheAsnGlyGlyLeuValValPheArgGlnThrGlnSer 696
DB 207 AATGAGGTGGTCTACTTCTCCGGGAAATCAAAATAGCGCTATAAAGCTTTCTCG 266
QY 697 SerSerGlySerLeuTyrPalaIleGlnGluThrValHisIleLysGlnGlyGluThrPhe 716
DB 267 GAACCT---TTCTTCCAGGAGAAAGAAACACATGATGCTCGCCCAAGGGAGAACCA 323
QY 717 ValTyrLysProValValGluGlyPro-----LlePro 727
DB 324 GATATGTCCTCCGAGGAGTACGAGCACTTCTCTCATATCAATGCCAGCTTCTTCC 383
QY 728 AspGlySerTyrIleArgAlaThrLeuHisAlaPheValAsnGlyGlnGlnGluLeuTyr--- 746
DB 384 GATGGCACCTATTACATCTGATCAAG-----GAGCAGGAGATTGTTGGAT 428
QY 747 -----LeuLysGlyLysArgAsnTyrThrValLysIleValAsnGlyThrAlaVal 763
DB 429 CCGATCGATTTGTTGGGAGCTATTAATTCGATCCGCTCATTCAGCATCTCTCT 488
QY 764 GluAlaIleGluSerGlu-----GluIleArgValPheProAsnProAlaArgAsp 781
DB 489 TCGGACATCCCTGTAGAGGATGTTTCTACTATAGTCTTATCCCAATCTGCTCAGCAG 548
QY 782 TyrValGluIleSerAlaProCysIleProGlnGluThrSerIleIleuPheAsnLeu 801
DB 549 TATGTCATGATACCATCTCTCCACATATGCGGAGCAGACATTCGTTGTTGATAT 608
QY 802 SerGlyLysIleValMetLysAsnSerLeuSerAlaGlyHisGlyArg 817
DB 609 CAAGGCGAATGACCTCTCCACGAAATGAAATGCGCATATGCGT 656

```

RESULT 8

US-09-453-702B-39
Sequence 39, Application US/09453702B
Patent No. 6365723

GENERAL INFORMATION:

APPLICANT: Blattner, Frederick R.
Burland, Nicole T.
Perna, Valerie
Plunkett, Guy
Welch, Rod

TITLE OF INVENTION: No. 6365723el Sequences of E. coli 0157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESS: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI

COUNTRY: US
ZIP: 53701-2113

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453, 702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110, 955
FILING DATE: 04-Dec-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296, 95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 25165
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-09-453-702B-39

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:
0.0017	145.00	25165	193
Percent Similarity:	34.78%	Conservative:	125
Best Local Similarity:	21.09%	Mismatch:	368
Query Match:	3.27%	Indels:	230
		Gaps:	48

US-10-030-330-1 (1-843) x US-09-453-702B-39 (1-25165)

```

QY 67 GlyIleThrSerGlnGluGlySerProAlaTyrPheTyrValAlaAsnArg-GlyAs 86
DB 2787 GCGGTAGCGGAGGAAATTCAGCGCGTGTGATGGGAAAGCCGACAGCAATAC 2846
QY 86 nasnglgllytrAlaLeuValAlaAspAspArgIleProThrIleLeuAlaTyrse 106
DB 2847 GGTACCGGCGCGGAGCTAATACGCTGATACCAACCGCAACG---CTTGCAATCA 2903
QY 106 rProIleGlyArgPheAspMetAspSerMetProAspAsnLeuArgMetTyrLeuGln 126
DB 2904 CACCATGCTCAG-----GACMACAT-----ATCAGTGC 2933
QY 126 eTyrAspGlnGluIleGlyLeuIleLeuSerGly-----LysAlaGlnLeuAsnGln 144
DB 2934 TCGGAGACATTAAGCTCGCTGTGATGACGCGGACGATGCAATGCGAAGGCGCAAC 2993
QY 144 uIleLeuArgThrGlnGlyValProAlaGluValHisAlaLeuMetAspAsnGlyHisPh 164
DB 2994 CGTAACACTGACCGTCAACGGAAGCCATACAGCAACCGTCTGAGGAGCGGA----- 3048
QY 164 eAlaAsnAspProMetArgTyrAsnGlnGlyTyrProTyrPheAsnHisLysGluProLeu 184
DB 3049 -----ACCTGGCACTGACCGCTGCTGCCAGGAGTCCAG---GCACT 3089
QY 184 uProAsnGlyAsnHisAlaTyrThrGlyCysValAla-----Thr-A 198
DB 3090 GCGGAGGAGTAAATTAAGCTGCAATGCCAGTGCAGCGATCGGAGGAGGACACCAAC 3149
QY 198 laAlaIleGlnIleMetArg-----TyrHisSerTyrProLeuGlnGlyGlyLys 215
DB 3150 CCACAGCGGAAATTCACGCTGACACCTACACCGCCCGTGG----- 3190
QY 215 eArPheAspTyrHisAlaGlySerLeuValGlyAsnTyrSerGlyThrPheGlyGlyMet 235
DB 3191 -----TCAGTGTAAATACCGTGGCGGCGGCGAT----- 3220
QY 235 yTrAspTyrIleAsnMetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnVal 255

```



```
Dh 3221 -----TTTTAAATGCGGAGCGCGTCG-----CCAGATCATCTCCGG 3263
Qy 255 spAlaTyrAlaThrLeuMetArgAspValSerAlaSerValSerMetSerPheTyr----- 273
Db 3264 ACAAGTCAACGGGTCTTCTCCAGGCGATACG--GTAAAGGTGAATTTGGGCACATCATGTCC 3322
Qy 273 ----- 273
Db 3323 TGACGGGCATCGTCTGGCAGATGCGAGCTGGAATGTGGCGCTGGACCCAGCGGTAAACC 3382
Qy 274 -----GluAsnGlySerGlyThrTyrSerValTyrValValGlyAlaLeuArgAsn 291
Db 3383 GCACGCTGGATCGCGAGCAATACGATTTTGTCTACCCCTGACAGATGGTGGCAGGAATA 3442
Qy 291 snPheArgTyrTyrAspSerLeuGlnLeu-----HisValArgAlaLeuTyrThr-----S 308
Db 3443 CTGGCGCGCGCTGTGAGCAATCAGCGTGTGGTCTTCTCCGTGTATCCACATTAAACA 3502
Qy 308 ergGlnGluTyrPheHisAspMetLeuArgGlyGluLeuAlaSerGlyATGProValTyrTyr 328
Db 3503 CCGTCTCCGCGCATGACATTATCAGTGC-----GCAAAAAAGGTGGCCACCTACCTTA 3559
Qy 328 IeGlyAsnAsnGlnSer-----IleGlyHis 337
Db 3560 CCGGTAGCACTCAACAGCTGAGACAGCAAAACGTCACAGTAACCTGTGGTGGACAGA 3619
Qy 337 IaPheValCysAspGlyTyrAlaSerAspGlyThrPheHisPheAsn-----Trp 354
Db 3620 GTTTTACCACTACCGCTGACGCC--GATGGCTCTGGAGTGTGACGGTACGCTGCCCGC 3676
Qy 354 IYTrGlyGlyValSerAsnGlyPheTyrTyrLeuThrLeu--LeuSerProThrSerL 373
Db 3677 CGATGGGAATATCGCTGACGCGCGCGGTGCGCATTCACCGCTGTGACAGGATCTACGC 3736
Qy 373 euGlyTyrIeGlyGlyGlnGlyIleGlyPheThrIle--TyrGlnGluIleIleThrGly 392
Db 3737 GCAAATACGCGCAACACTCCCGCACCATTAACCGTCATAGACAGCGCCGCGCTTAAGCA 3796
Qy 392 IeGluProAlaLys-----ThrProAlaGluAlaGlyThrAspAlaLeuP 407
Db 3797 TTGATTCACCTGACCGCTGATTAACATCAATTAACGCCCGCAAAAGCGGAGAGAT--CTGC 3853
Qy 407 roIleLeuAlaLeuLysAspIleGluAlaGlyTyrLysSerGlnSerGlyLeuAsn---- 425
Db 3854 CCATACCGCGCACACCGACGCGTCAAGCGCGGCGACAGCGTGACCTTACGTAAATGGGC 3913
Qy 426 -----ValGlyTyrSerIleTyr 432
Db 3914 AGACGTATCAGGCGCTGTCAGCCAGACGCGACCTGAGCGTGTGCGCGCGCGCA 3973
Qy 432 snThrGlyGluLeuGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGly 452
Db 3974 ACGTGGGCGCA-----CTGGCGAGCGCA 3997
Qy 452 luValIleGluValLysThrSerSerIleAsnIleSerTrpTyrGlyTyrGluHisP 472
Db 3998 ACGCT-----ACGGTCAACCGCGCAGCGTAAAC-----GATGTCCCGCGGTAAATC 4039
Qy 472 roGluSerPheSer-----LeuAlaP 479
Db 4040 CCAGACAGCGTTTACGCGGTGGCGGTGTGATGCCACGCGCGCGGTGTAAACATTAAAC 4099
Qy 479 roAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuLeuTyrArgLys-----Thg 497
Db 4100 CCGTGGCGCACCAATAACGTATTCATCCCGCAACATCTCAGCGCGCAATCATCATCAG 4159
Qy 497 IYThrGluGlnTrpGluProValArgHisAlaGlnGlyGlyTyrValAsnSerIleLys 517
Db 4160 GCACG-----GTACGCGCGCTCAGCGCGCGGTATTCGTCACCGTACGC 4204
Qy 517 alAsnThrThrAspProAsnAsnValValValThrValAspAsnGlnGlyLysLeuS 537
Db ----- 537
```

```
Dh 4205 TGATAATATGTGAT-----TACACCAAGGTGTGATGTTCCGCAACTGCA 4252
Qy 537 erIle---ValProAsnSerPheValAlaAspLeu-----AsnSerTyrGluHisSer- 553
Db 4253 GTCTGGCGTTCGCGCTCGTGTGTGTAGTGGGTGGCGGACGCGCATTAATCTGTCTACGC 4312
Qy 554 -----ThrIleThrValGlnPheAsn 561
Db 4313 TCTCGGTAAACCGCAAAAGCGGAAACACGCGGACCGCATATGACCGTACCGTCAATA 4372
Qy 561 erAspSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrGlyAlaThr 581
Db 4373 CCGCGCGGCC-----CTTATCGCATTAACAGCATTTGCG--G 4408
Qy 581 IaAspAspValIleSerLeuGlyTyrValMetAlaGluValPro---GlyGlySerSer 600
Db 4409 GCAGTATGTGATTAACCGCAGGCAAAAGGCGCGATTCACAGATTACCGCAGCGC 4468
Qy 600 snTyrProValTyrPheSerLysAspValLeuThrLeuSerGlnGlyAspTyrThrLeu 620
Db 4469 ATCAGCGCTTAAACCGCCATCAGCGT---ACGCTGAACGGCAAAATTTACAC--- 4521
Qy 620 rPTyrArgPheSerIleAsnAsnGlnLysAspGluTyrLysIleGlySerValSer 640
Db 4522 -----ACCAGCAGCGACGCTCCGCAACTG-----ACG 4552
Qy 640 alLysThrProThrGluTyrThrHisProLeuPheGluValGlyHisAsnGlnThrSer 660
Db 4553 TCACCGTTCGCGCATCGCGGCTTACACAGATTAGCCGAGCCAACTAATACGTAACGCGCG 4612
Qy 660 hrTyrThrLeuAspMet-----AlaHisAsnArgValLeuProAspPhe 675
Db 4613 CCGTGAACACAGCATATCGGACAGACGCGCCACTGCAACCATTAACGTCTGTGTGACAGCG 4672
Qy 675 hrLeuLysAsnLeuGlyLeu---ProPheAsnGlnGluLeuValValPheArgLys- 693
Db 4673 CGCTGCCCGGTGTGACCATTAATCCGTGTGCAACCGCATTAATTAACGCGCGCGCA 4732
Qy 694 -----ThrGlnSerSerGlySerLeuThrAlaAlaGln-----GluThrValH 709
Db 4733 CCGCGCGTGGCGCAACCATTCAGCGCGGCGCATGTGCGGAGGAGATGGCGACAGCTAA 4792
Qy 709 IsIleLysGlnGlyGluThrPheValTyrLysProVal-----V 722
Db 4793 CTATTAAGTGGGTGTGATTAATCTTATACGCGACGCGTGTGGCGCAATCTCACTGAGCG 4852
Qy 722 alGluGlyProIleProAsp-----GlySerTyrArgAlaThrLeuHisAlaP 738
Db 4853 TGAGCGTTCACAGCGGACAGATATTCAGCGCTGTGGAAATGTGCGATTAAATGCTT 4912
Qy 738 heValAsnGlnGlnGlnLeuThrLeuLysGlyLysArgAsnTyrThrValLysIle- 757
Db 4913 CAGTCAACCAATTAACAGCGCACCGCGCAGCGCGGAGATTCACCATCGACCGCA 4972
Qy 758 --ValAsnGlyThrAlaValGluAlaIleGluSerSerGlnGluIleArgValPhePro 777
Db 4973 ATGTGCCCGCGCTCGCGGTGATACGCTGCGCGGCGCATGATGTGTGAATATCATGAGC 5032
Qy 777 snProAlaArgAspTyrValGluIleSerAlaProCysIleProGlnGluThrSerIle 797
Db 5033 ACGGAGAGCGGTGTGATCGCGCAGCAGCTCGGGGTGGCTGAAGACACCGCGCTTA 5092
Qy 797 IeLeu-----PheAspLeuSerIleLysIleValMetLysAsnSerLeuSer 813
Db 5093 CCGTTACGATTAATATGTGGAATTAACACCTGCGGTGAGCGCGCATGTAGTGTGAGCG 5152
Qy 813 IaGlyHisGlyArgMetAspValSerArgLeuProAsnGlnGlyAlaTyrIleLeuLysVal 833
Db 5153 TGGCGCTCAAGCGCGCGCAGGTATACGCCCTCGCGCTCGGGAGCGGATTAATATGCGCTT 5212
Qy 833 spGlyTyrThrThr-----LysIleAsnIleValHis 843
Db 5213 CAGGGGAAGTACGCGCGCAACTCGGTGAGCATTAACGAT 5253
```

RESULT 9

US-08-119-125A-3

Sequence 3. Application US/08119125A

Patent No. 5610011

GENERAL INFORMATION:

APPLICANT: SMITH, Hilda Elizabeth

APPLICANT: VECCH, Uri

TITLE OF INVENTION: DNA sequences which code for virulence

TITLE OF INVENTION: Characteristics of Streptococcus suis and parts thereof, poly

TITLE OF INVENTION: Antibodies derived therefrom and the use thereof for the diag

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESSES:

ADDRESSER: Central Diergeneeskundig Instituut

STREET: Edelhertweg 15

CITY: PH Leijst

STATE: The Netherlands

ZIP: NL-8219

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage

OPERATING SYSTEM: MS-DOS V.6.0

SOFTWARE: Wordperfect V. 6.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/119,125A

FILING DATE: 20-SEP-1993

CLASSIFICATION: 514

APPLICATION NUMBER: PCT/NL92/00054

FILING DATE: 19-MAR-1992

APPLICATION NUMBER: NL 9100510

FILING DATE: 21-MAR-1991

ATTORNEY/AGENT INFORMATION:

NAME: Handal, Anthony H.

REGISTRATION NUMBER: 26275

REFERENCE/DOCKET NUMBER: SMITHHE119125

TELECOMMUNICATION INFORMATION:

TELEPHONE: (203) 838-8589

TELEFAX: (203) 838-8794

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 4118 base pairs

TYPE: Nucleic acid with corresponding amino acids

STRANDEDNESS: single stranded

TOPOLOGY: linear

MOLECULE TYPE: genomic DNA

ORIGINAL SOURCE:

ORGANISM: Streptococcus suis type II (pathogenic)

FEATURE:

FEATURE: NAME/KEY: proline rich region

LOCATION: from bp 2757 to 3014

FEATURE: NAME/KEY: repetitive units

LOCATION: from bp 3015 to 3176, 3423 to 3584 and 3585 to 3743

FEATURE: NAME/KEY: membrane anchor sequence

LOCATION: from bp 3825 to 3926

FEATURE: NAME/KEY: dyad symmetry regions

LOCATION: from bp 4069 to 4080 and from bp 4087 to 4098

US-08-119-125A-3

US-10-030-330-1 (1-843) x US-08-119-125A-3 (1-4118)

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-10-030-330-1 (1-843) x US-08-119-125A-3 (1-4118)

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

FEATURE: NAME/KEY: proline rich region

LOCATION: from bp 2757 to 3014

FEATURE: NAME/KEY: repetitive units

LOCATION: from bp 3015 to 3176, 3423 to 3584 and 3585 to 3743

FEATURE: NAME/KEY: membrane anchor sequence

LOCATION: from bp 3825 to 3926

FEATURE: NAME/KEY: dyad symmetry regions

LOCATION: from bp 4069 to 4080 and from bp 4087 to 4098

US-08-119-125A-3

US-10-030-330-1 (1-843) x US-08-119-125A-3 (1-4118)

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-10-030-330-1 (1-843) x US-08-119-125A-3 (1-4118)

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

1047 GTGTTAGGAATAT-----ACGATGACCGGCGAGGTGTTACGTTTTCACATCCG 1100
OY 245 AspleuaspasnLeuThrGlnSerGlnValasp---AlaTyrAlaThrLeuMetArgasp 263
Db 1101 ACMAAGTAATGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCTCAAT 1160
OY 264 ValSerAlaSerValSerMetSerPheTyrGlnAsnGlySerGlyTyrSerValTyr 283
Db 1161 GCTATTAATGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCTCAAT 1217
OY 284 ValValGlyAlaLeuArg---AsnAsnPheArgTyr----- 294
Db 1218 ATGTAATGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCTCAAT 1277
OY 295 ---LysArgSerLeuGlnLeu-----HisValArgAla 304
Db 1278 GCACCTTAATGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCTCAAT 1334
OY 305 LeuTyrThrSerGlnGlnLeuTyr-----HisAspMetLeuArgGlyGlnLeuAla 320
Db 1335 GTATTAATGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCTCAAT 1394
OY 321 SerGlyArgProValTyrTyrAlaGly----- 329
Db 1395 AAGGA---ACGTAATGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCT 1451
OY 330 ---AsnAsnGlnSerIleGlyHisAlaPheValCysAspGlyTyr 343
Db 1452 ATGCTGAATTAATGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCT 1496
OY 344 AlaSerAspGlyThrPheHisPheAsnTrpGlyTyrGlyVal---SerAsnGlyPhe 362
Db 1497 TATTAATGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCTCAAT 1556
OY 363 TyrIleLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyGlyIleGlyPhe 382
Db 1557 ATGAGTGTGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCTCAAT 1598
OY 383 ThrIleTyrGlnGlnIleIleIleThrGly-----IleGluProAla 395
Db 1599 ACTATTAATGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCTCAAT 1658
OY 396 LysThr----- 397
Db 1659 AAGTAATGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCTCAAT 1718
OY 397 --- 397
Db 1719 ATGCTCAATGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCTCAAT 1778
OY 398 ---ProAlaGlnAlaGlyThr----- 403
Db 1779 ACTGAGTGTGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCTCAAT 1838
OY 403 --- 403
Db 1839 GACCAAGTGTGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCTCAAT 1898
OY 404 ---AspAlaLeuProIleLeuAlaLeuLys----- 412
Db 1899 CTTGCAATGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCTCAAT 1958
OY 413 ---AspIleGlnAlaGlnIleTyrLysSerGlyLeuAsnValGlyTyrSer 429
Db 1959 AAGTAATGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCTCAAT 2018
OY 430 IleTyrAsnThrGlyGlnGlu-----GlnSerAsnLeuAspLeuGlyTyrArg 445
Db 2019 GTTGAATGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCTCAAT 2078
OY 446 ---LeuAsnLysAlaAspGlyGlnVal----- 453
Db 2079 CCAGCAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2138
OY 454 ---IleGluVal 456
Db 2139 TCAGTAAGACAACCGGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 2198
OY 457 LysThrSerSerIleAsnIleSerTrp----- 465
Db 2199 AAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2258
OY 466 ---TyrGlyTyr----- 468
Db 2259 GTTTCAGTAAGACAACCGGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 2318
OY 469 ---GlyGlnHisPro 472
Db 2319 TTAGCTTCATGCTTTGTAAGACAAGTAATCAATTTGCTGCTGCTGCTGCTGCTGCTGCT 2378
OY 473 ---GlnSer 474
Db 2379 GTTGTGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2438
OY 475 PheSerLeuAlaProAsnGlnLeuSerGlnGlyIleAsn---ThrIleThrLeuTyr 493
Db 2439 ATTGAACCAACAACAGGCAAAATGTAAGCGGCTGTAACAAAGAAATTAATGCTGCTAT 2498
OY 494 ArgArgThrGlyThrGlnGlnIleProValArgHisAlaGlnGlyTyrValAsn 513
Db 2499 ACAGCAGTACAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2558
OY 514 SerIleLysValAsn-----ThrIleLysPro 522
Db 2559 GATCCAGAAACGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2618
OY 523 Asn---AsnValValValThrValAspAsnAsn----- 532
Db 2619 AATCAACCAAGTAATGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCT 2678
OY 533 ---GluGlyLysLeuSerIle----- 538
Db 2679 GATGTGAGAAATGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCTCA 2738
OY 539 ---ValProAsnSerPheValAlaAspLeuAsnSerTyrGlnHisSer 553
Db 2739 AATGTTGCAAACTGATTCAGAG-----ATTCCAAATGTACCAAAACAGCCCTCA 2792
OY 554 ThrIleThrValGlnPheAsnSerAspSerProAspGlnIleArgThrProValAlaPhe 573
Db 2793 AAGTAATGCTCAATTAATGCAATTAATGCTCAATTAATGCTCAATTAATGCTCAAT 2843
OY 574 AlaLeuSerThrGlyAlaThrAlaAspValIleSerLeuGlyTyrPylMetAlaGlu 593
Db 2844 ---ACGACACCAAGCAAAATGCGAGGTTCCAAATATTCCTTAC----- 2885
OY 594 ValProGlyLysSerSerAsnTyrProValValIleThrSerLysAspValLeuThrLeuSer 613
Db 2886 GTTCCAGCA-----TATACACCGGTTGATCTTAAGATTAACACGCCGTTGAAA 2933
OY 614 GluGlyAspTyrThrLeuTyrPylArgPheSerIleAsnAsnGlnLysAspGlnTyrLys 633
Db 2934 CCAATTTGAT-----CCAATATGATCAGGTAAAG 2960
OY 634 LysIleGlySerVal-----SerValLysThrProThrGlnTyr 646
Db 2961 ---GGTTATGTACCAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 3014
OY 647 ThrHisProLeuPheGlnValGlyHisAsnGlnThrSerThrTyrThrLeuAspMetAla 666
Db 3015 GTT---CCAGTTAAAAAATGCTGATTAACCAAC-----GTTATGTAAGAG 3056
OY 667 HisAsnArgValLeuPro-----AspPheThrLeuLysAsnLeuGlyLeuPro----- 682
Db 3057 GGTAAACCTATGCAACCGCAAGAGGCAACCAAAACCAAAACCAAAACCAAAACCAAAACCA 3116

276 -----glyserglythr----- 279
Db TACATGACGCCGTTACGCCAAATTAGGTGGCGTACCGCTTTTCTCTCTGCA 875
Oy 280 -----TyrSerValTyrVal----- 284
Db AACATGTTGGCTATTAACGCTCTTCATGATCAGATTTTCTGATATACCGCTTA 935
Oy 285 -----ValGlyAlaLeuAlaGlyAsnAspPheArgTyrLysArgSerLeuGlnLeuHisValArg 303
Db GGTATGTTGGCGAATAGCTGCGAGACTATTTCAAAGTAGC-----GTAAAC 983
Oy 304 AlaLeuTyrThrSerGlnLutPheHisAspMetLeuArgGlyGluLeuAlaSerGlyArg 323
Db GGCATATTCCGCATGAGCGCGTGCATGATCATCCATACAGAAAGACTATGATAGCGC 1043
Oy 324 ProValTyr-----TyrAlaGlyAsnAsnGlnSerIleGly 335
Db CCAGCAATGGCTTCGATATCCGTTTAAAGGCTATACCGCTCATATCCGCAATTAGGC 1103
Oy 336 HisAlaPheValCysAspGlyTyrAlaSerArgGlyThrPheHisPheAsnTrpGlyTrp 355
Db GCCAAGCTGATATATGACAGCATATTATGATGATTAATGTTGCTTTTAAATCTGAT--- 1160
Oy 356 GlyGlyValSerAsnGlyPheTyrLysLeuThrLeuLeuSerProThrSerLeuGlyIle 375
Db 1161 -----AACTG-----CACTCCAAATCTGTGTCG 1184
Oy 376 GlyGlyGlnGlyIleGlyPheThrIleTyrGlnGlnIleIleThrGlyIleGluProAla 395
Db 1185 GCGACCGTGGTGAACCTATACCCGATCTCTGTCGATGAGGATC----- 1235
Oy 396 LysThrProAlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGlu 415
Db 1235 ----- 1235
Oy 416 AlaGluTyrLysSerGluSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGlyGlu 435
Db 1236 -----GATTAACCT-----CATGGTAACGGTAAAT 1259
Oy 436 GluGlnSer-----AsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyGlu 452
Db 1260 GAAATGATCTCTCTTACTCTATGACAGTCCGTTATCAGTTGATTAATCCGCGTCCAG 1319
Oy 453 ValIleGluValLysThrSerSerIleAsnIleSerTrpGlyTyrGlyGlnHisPro 472
Db 1320 CAAATTTGAACACAG-----TATGTTAAACGAGTTA 1349
Oy 473 GluSerPheSerLeuAlaProAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuLeu 492
Db 1350 AGAACATTATTCAGCAGCGCTTACAGTGTGTTGAGCTTAATACAAATATTTTGTGAG 1409
Oy 493 TyrArgArg-----ThrGlyThrGlnGln 500
Db 1410 TACAAGAAGCAGATATTCTTCTCTGAATATCCGATGATATTAAAGTACGACAGC 1469
Oy 501 -----Trp 501
Db 1470 AGTACGCAAGAAAGTCACTGATGCTTAAGCAAAATAGCGTCTGATGCTATGCTGCG 1529
Oy 502 ---GluProValArgHisAlaGlnGlyTyrValAsnSerIleLysValAsnThrThr 520
Db 1530 GATGATAGTGCATTAACGAGTACAGCGCGTCAAGATTCAGCATAGCGGAACCAAGCGCA 1589
Oy 521 AspProAsnAsnValValValThrValAspAsnAsnGlnGlyLysLeuSerIleValPro 540
Db 1590 CAAGACTACAGGCTATTTCTGCTGCTATGTCGAAGGTGCGAGC----- 1634
Oy 541 AsnSerPheValAlaAspLeuAsnSerTyrGlnHisSer----- 553
Db 1635 AATATTATTAAGAATGACGCTCGCGCTATGACCGTAAATGCAATAGCTTAACAATGTA 1694

Oy 554 -----ThrIleThrValGlnPheAsnSerAspSerProAspGluIleArgThrProVal 571
Db 1695 CAGCTTACATTAACCGCTTGTGCGAATGTCGATGTCGACACGAGTGGGTAACGGAC 1754
Oy 572 AlaPheAlaLeuSerThrGlyValIleThrAlaAspAspValIleSerLeuGlyTrpValMet 591
Db 1755 TTACGCGGATTAAGACTTCGCGTAAAGCGGATTAACCGCATACCATTTACTATACCGCG 1814
Oy 592 AlaGluValProGlyGly-----SerSerAsnTyrProValValTrpSerLysAspVal 609
Db 1815 ACGGTGAAGAAAGATGGTACCTCAGCGTAAATGTCCTGTTCAATTAAT----- 1865
Oy 610 LeuThrLeuSerGlnGlyAspTyrThrLeuTrpTyrArgPheSerIleAsnAsnGlnLys 629
Db 1866 -----ATTGTTCAAGACCTGCACTT-----GGGCAATTAAGTCCCAA 1907
Oy 630 AspGluTrpLysLysIleGlySerValSerValLysThrProThrGlyTyrThrHisPro 649
Db 1908 ACGGATGCTAAACGGTAAGCAACCGTAACTGTAAGTCAAGTACGCCA----- 1955
Oy 650 LeuPheGluValGlyHisAsnGlnThrSerThrTyrThrLeuAspMetLanHisAsnArg 669
Db 1956 -----GACAGCTGCTGCTGCTGCTAAACCGCGAGATGAGTCA----- 1997
Oy 670 ValLeuProAspPheThrLeuLysAsnLeuGlyLeuProPheAsnGlyGluLeuValVal 689
Db 1998 -----GCACTTAATGCCAGTGGGTTAA 2021
Oy 690 ValPheArgGlnThrGlnSerSerGlySerLeuThrAlaAlaGlnLutThrValHis 709
Db 2022 TTTTGTGATCAACCAACGACGATTAAGCTGATGAT---AAGCGTATTAAGCAACTGCA 2078
Oy 710 IleLysGlnGlyGlu-----ThrPheValTyrLysProValValGluGly---Pro 725
Db 2079 GTAGCAATGTTGAAGATGCTTAATTAATCTGTAAGTAAGTAAAGGTAAGCGTACGCA 2138
Oy 726 IleProAspGlySerTyrArg---AlaThrLeuHisAlaPheValAsnGlyGlnGln 744
Db 2139 GTTAATTAATCAATCCGTTATCATTTCAACAACTTGGGATGTCAACGGTAAGTCA 2198
Oy 745 LeuTyrLeu-----Lys 748
Db 2199 ACGCAAGCAACCAACGGAATGATGTCGTCGACGATTAACACTTCCAGTTCGCGC 2258
Oy 749 GlyLysArgAsnTyrThrValLysIleValAsnGlyThrAlaValAlaGluAlaGlnSer 768
Db 2259 GGTAAAGCGACTGTTAGTGCACAGTCACTGATGAGGCTGAGGTTAAACGACTAGGTC 2318
Oy 769 Ser-----GluGlnIleArgVal----- 774
Db 2319 ACTTTTGTGATGACGTAATAATGTCGCAATATGTCGATTAACGTAAGCAAGCGGTGCT 2438
Oy 775 -----PheProAsnProAlaArgAspTyrValGluIleSerAlaProCys----- 789
Db 2379 GCGAGATGCGCTAAATATTGCGTCAATATGTCGATTAACGTAAGCAAGCGGTGCT 2438
Oy 790 -----IleProGlnGlnThrSerIleIleLeuPheAsnSerGly 803
Db 2439 GATGTAATATTCATGATGATTCAGAAATATACCAAAATGCGCACTGTCTCATACATCAGG 2498
Oy 804 LysIleValMetLysAsnSerLeuSerAlaGlyHisGlyArgMetAspValSerArgLeu 823
Db 2499 AAAGTCACTTGAAT-----GTTAAAGCAGAGTCT---GTAATTAAGCC 2540
Oy 824 ProAsnGly-----AlaTyrIleLeuLysValAspGlyTyrThrThrLys 838
Db 2541 ACATCTGGGATTAAGCAACAGTAAAGTATACATAAAAGCAACCGCTATATGATATAAA 2600
Oy 839 IleAsn 840
Db 2601 GTGAT 2606

RESULT 11

US-09-696-188B-21
Sequence 21, Application us/09696188B
Patent No. 6406885
GENERAL INFORMATION:
APPLICANT: Stewart, C. Neal
McKee, Marian L.
O'Brien, Allison D.
Machtel, Marian R.
TITLE OF INVENTION: Method of stimulating an immune response
By Administration of Host Organisms That Express Influenza
Antigens.
Alone Or As A Fusion Protein With One Or More Other
Antigens.
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flinnegan, Henderson, Farabow, Garrett &
Dunner, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/696,188B
FILING DATE: 26-Oct-2000
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/840,466
FILING DATE: 1997-04-18
ATTORNEY/AGENT INFORMATION:
NAME: Boone, Laura L.
REGISTRATION NUMBER: 43,505
REFERENCE/DOCKET NUMBER: 04955-0029-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4400
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 3106 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 21:
-696-188B-21
Alignment Scores:
Seq. No.: 0.000186 Length: 3106
Score: 140.00 Matches: 167
Percent Similarity: 30.67% Conservative: 128
Best Local Similarity: 17.36% Mismatches: 331
Query Match: 3.15% Indels: 336
Gaps: 45
US-10-030-330-1 (1-843) x US-09-696-188B-21 (1-3106)
Oy 74 glyserproalatyphetyrvalalaasnarg-----glyasnanglucltytyrvala 91
Db 144 GGTTCAGCATGTTGTTTATGTTAATCAGAAATTCATTGCAAAATGTAATTTATTT 203
Oy 92 leuvalalaalaaspargilperthrilleuvalatyrseryproillelytyrphe 111
Db 204 AAATGGGTTCGATCAAACTGTTACTCATGATACATCAG-----AATGCCCT 257
Oy 112 -----Aspmetaspsersermetproasp---asn 119
Db 258 TTTTATACGTTGAAACATGTCGAACGTCGTCATCTTTCAAAATCCCAAGATTTTAT 317
Oy 120 leu---argmettrp-----leucinilleytraspngluilleglyleuileu 135

Db 318 TTATCGACGATTTGGCTGTCGTAATAGCATTTATACATTCGTAAGAAAGCAATGATGAAG 377
Oy 136 SerGlyLysAlaGlnLeuAsnGluGluLeuAlaGlyThrGlnGlyValProAlaGluVal 155
Db 378 GCCGGCGCTGGTCAG-----CAGATTCATTTGGCCACGTCACCAAAATCCCTTTAATAC 431
Oy 156 HisAlaLeuMetAspAsnGlyHisIspheAlaAsnAspPrometArgTrpAsnGlnGly 175
Db 432 AGTCGACAT----- 440
Oy 176 ProTrpAsnAsnLysGluProLeuLeuProAsnGly----- 187
Db 441 CCACCTTTAGGTTCCGACACCTCTTCTGTCGAGCGTGCTGTGTGTCACAGCAATAAA 500
Oy 188 -----AsnHisAlaLeuThr 192
Db 501 CTGACATAAATGTCCTCCGAGCTGACCAAAACCAATGACCGATGACAGCATTAAT 560
Oy 193 GlyCysValAlaIleThrAlaAlaGlnIleMetArgTyHisSerTrpProLeuGlnGly 212
Db 561 TATCGCGCAACAACAGCGCGGCGAGCTCGGTAGCCAGCTTCAGTCGATCCGTGTAAGAGC 620
Oy 213 GluGlySerPheAspTrpHisAlaGlySerLeuValGlyAsnTrpSerGlyThrPheGly 232
Db 621 GATTACGGCAAGATACCGCTCTTGTG---ATCGCTGTAAACCAAGCTTCGCA----- 671
Oy 233 GluMetTyTrpAspTrpIleAsnMetProGlyAsnProAspLeu-----Asp 247
Db 672 CAGTTGCAAGCGCTGTTCAACATTTATGCAAGCGAGTTAATCTGCAGAGTGAT 731
Oy 248 AsnLeuThrGlnSerGlnValAspAlaTyAlaThrIleMetArgAspValSerAlaSer 267
Db 722 AACTTTGACGAGTACCTGAC----- 755
Oy 268 ValSerMetSerPheTyGluAsn----- 275
Db 756 TTTTATTAACGTTCTATGATTCGAAAAAATGCTGCGATTTGTCAGTCGAGCGCGCT 815
Oy 276 -----GlySerGlyThr----- 279
Db 816 TACATTGACCTCCGCTTACGCGCAATTTAGTGCGGGGTGACGCTTTTCTCTCTGCA 875
Oy 280 -----TySerValTyVal----- 284
Db 876 AACATGTTGGCTATACGCTTTCATGATCAGATTTTCTGCTGTAATACCGCTTGA 935
Oy 285 ---ValGlyAlaLeuAlaArgAsnAsnPheArgTyLysArgSerLeuGlnHisValArg 303
Db 936 GGTATTGGTGGCGAATTAAGTGGCGAGACTAATTCGAAAGTAGC-----GTAAAC 983
Oy 304 AlaLeuTyTrpHisSerGlnIleThrHisAspMetIleArgGlyLeuAlaSerGlyArg 323
Db 984 GCGATTTTCCGATGAGCGCGGATGATGATCATCAATGAAAGCATGATGATGAGCGC 1043
Oy 324 ProValTyTrp-----TyAlaGlyAsnAsnGlnSerIleGly 335
Db 1044 CCAGCAAAATGCGCTGATATCCGTTTAAATGCGCTATACCGCTATACCGGCAATTAAGC 1103
Oy 336 HisAlaPheValCysAspGlyTyTrpAlaSerAspGlyThrPheHisPheAsnTrpIleTrp 355
Db 1104 GCCAAGCTGATATGATGACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1160
Oy 356 GlyGlyValSerAsnGlyPheTyTrpLysLeuThrLeuLeuSerProThrSerLeuLys 375
Db 1161 -----AAGCTG-----CAATCGAATCCGTGTCG 1184
Oy 376 GlyGlyGlnGlyIleGlyPheThrIleTyGlnGlnIleIleThrGlyIleGluProAla 395
Db 1185 GCGACCGTTGGTGAATATACATCCGATTCCTGTCGAGCATGGGATC----- 1235
Oy 396 LysThrProAlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGlu 415

Db 1235 ----- 1235
Qy 416 Alagutrylserglyleuasnvalglytyrseriletyrasnthrglylu 435
Db 1236 ----GATTACCGT-----CATGGTACGGGTAAT 1259
Qy 436 Gluglnser-----Asnleuaspleuglytyrtyrleuasnlyalaaspglylu 452
Db 1260 GAAATGATCTCTTACTCAATGCACTCCGTTATGATGTAATGATGCTGTCAG 1319
Qy 453 Valilegluvallysthrserleasnilesertpyrlytyrlygluinhispro 472
Db 1320 CAATATGACACAG-----TATGTTAACGAGTTA 1349
Qy 473 Gluserheserleuaprosnleuarglnleuarglnlyleasntrilethleu 492
Db 1350 AGACATTTATCAGGACCGCTTACGATCTGTTAGGTAATACAAATATTATTCGAG 1409
493 TYRATGATG-----Thrglythrlyglu 500
1410 TACAGAGACAGATATTCTTCTGTAATATCCGATGATATTAATGACTGAACAC 1469
Qy 501 -----Trp 501
Db 1470 AGTACGAGAGATTCAGTTGATCGTTAAGACAAATACGCTCGATCGTCTGG 1529
Qy 502 ----Gluprovalarghisalaglytyrtyrvalasnserileysvalasnthr 520
Db 1530 GATGATRGTCATTTACCGACTCAGGGGCTGACATTCAGATACGAGACCAAGGCCA 1589
Qy 521 AspProsnasnvalvalthrvalaspasnleuarglnlyleuasnleuvalpro 540
Db 1590 CAAGACTACAGGCTATTCTGCTGCTATGTGCAAGCTGGCAGC----- 1654
Qy 541 AsnSerPhevalalaaspleuasnsertyrlythliser----- 553
Db 1635 AATATTTATTAAGTACGCGCTCGCTATGACCGTAAATGCAATAGCTTAACAATGTA 1694
Qy 554 ----Thrlthrvalglupheasnseraspserprospglnleaglythrproval 571
Db 1695 CACCTACTATTTACCGCTCTGTCGTAAGTGTCAAGTGTCAACGAGTGGGTAACGAC 1754
Qy 572 Alaphealaleuaserthrlyalathralaspaspyallilesertleuglytyrvalmet 591
Db 1755 TTACGGCGGATTAAGACTTGGCTAAGCGGATTAAGCCCATACCTACTTATACCGCG 1814
Qy 592 Alagluvalproglyly-----SerSerasntrprovalvaltyrserlyaspval 609
Db 1815 ACGGTGAAAAAGATGGGTGCTCAAGCTAATGCTCCCTGTTCAATTAAT----- 1865
610 leuthleuarglnlyasptrythrleuthrlytyrtyrgheserleasnasnleu 629
Db 1866 ----ATTGTTTCAGAGACTGCAACTCT-----GGGCAAAATAGGCCAAA 1907
Qy 630 AspGlntrplyslyleuglyservalservallysthrprothrlytyrthrhispro 649
Db 1908 ACGGATGCTAAGGTAAGCAACCGTACCTGAAGCGATACGCCA----- 1955
Qy 650 leuphegluvallyhisasnleuarglnthrserthrlythrleuaspmetalahisasnarg 669
Db 1956 ----GGACAGGTCTGCTGTCGTTAAACCGCGAGATGAGATGCTCA----- 1997
Qy 670 Valleuproasphehrleuylsasnleuglyleuproheasnleuglyleuvalval 689
Db 1998 ----GCACCTAATGCCAGCTGGCGTATA 2021
Qy 690 Valphearglnthrghisnserserglyserleuthrpalalaglnlythrvalhis 709
Db 2022 TTTTGTGATCAACCAAGCGACATTTACTGAGAT---AAGCTGATAGACAACTGCA 2078
Qy 710 Ilelysglnlyglu-----ThrPhevaltyrlyspvalvalglugly---Pro 725
Db 2079 GTAGCAAAATGCTAAGATGCTATTAATATACGTAAAGTATGAAAAACGGTCACCA 2138

Qy 726 Ileproaspglysertryrarg---Alathrleuhsalaphavalasnlyglngln 744
Db 2139 GTTATATATCAATCCGTTACATCTCTCAACAACTTTGGATGTTCAACGGTAAGTCA 2198
Qy 745 leuthleu-----Lys 748
Db 2199 ACGCAACCAACACAGGGAATGATGCTGCGACGATTAACACTACTTCCAGTTCCGCC 2258
Qy 749 Glylysarpsantyrthrvallysilevalasnlythrvalalaglualegluser 768
Db 2259 GGTAAACGCACTGTATGCGACAGTCAGTATGCGGCTGAGTTAAAGCGACTGAGTCC 2318
Qy 769 Ser-----GluGlnleargval----- 774
Db 2319 ACTTTTGTGATGAGTGAATGAAATGACACAAAGTTGATTTATGTTACATATGTCAGA 2378
Qy 775 ----Pheproasnproalaargasptryrvalgluileseralaprolys----- 789
Db 2379 GCGGATGGCTATATTTGGCTGCATATATGCTGATTAACGAAAGCAAGCGGTGCT 2438
Qy 790 -----IleproglngluThrserleileuaspheaspleserly 803
Db 2439 GATGTCATATTCATGATGATTCAGAAATACCGATACGCGACTGTCATGATCAGG 2498
Qy 804 Lysilevalmetlysasnserleueralaglyhisgltyrmetaspvalserargleu 823
Db 2499 AAGTCACTTTGAAT-----GTTAAAGCAAGTGTG---GTTAATTAAGCC 2540
Qy 824 Proasnly-----Alatyrilleuylsvalaspglytyrthrly 838
Db 2541 ACATCTGCTGATTAAGCAACAGTAAGTATACCTAATAAAGCACCGCTGATATGATAAA 2600
Qy 839 Ileasn 840
Db 2601 GTGGAT 2606

RESULT 12
US-08-277-231A-1
Sequence 1, Application US/08277231A
Patent No. 5643725
GENERAL INFORMATION:
APPLICANT: Green, Bruce A.
APPLICANT: Brinton, Charles C.
TITLE OF INVENTION: Sequence and Analysis of LKP p11n
Patent No. 5643725
TITLE OF INVENTION: Structural Genes and The LKP p11 Operon of No. 5643725tpa
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277, 231A
FILING DATE: 19-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.
REGISTRATION NUMBER: 32,542
REFERENCE/DOCKET NUMBER: ACC94-02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 1:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 9432 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: complement (1882..2532)
FEATURE:
NAME/KEY: CDS
LOCATION: 2854..3630
FEATURE:
NAME/KEY: CDS
LOCATION: 4016..6238
FEATURE:
NAME/KEY: CDS
LOCATION: 6259..6873
FEATURE:
NAME/KEY: CDS
LOCATION: 6955..8265
FEATURE:
NAME/KEY: CDS
LOCATION: 8395..9340
US-08-277-231A-1

Alignment Scores:
Pred. No.: 0.00115
Score: 140.00
Percent Similarity: 33.15%
Best Local Similarity: 21.24%
Query Match: 1
Matches: 196
Conservative: 110
MisMatches: 326
Indels: 291
Gaps: 46

US-10-030-330-1 (1-843) x US-08-277-231A-1 (1-9432)
QY 22 AlaProValThrLeuGlnArgAlaLeuSerLeuAlaArgLeuAlaLeuArg----- 38
DB 4571 GCCCGAGTGTGGCGTGGCGAACAACAAACGCCAAAGTCAACAAACAAATGCG 4630
QY 39 -----GlnValSerLeuArgMetGlyGlnThrAlaValSerAspLysIleSer 54
DB 4631 TATACGATTATTCATACCGCTCCCGAGGCGCTTCGTGATTAACGATTGTGATGCC 4690
QY 55 IleAspTyrValTyrArgGlnGlyAspAlaGluArgGlyIleThrSerGlnGluGlu 74
DB 4691 ACCGGTTAT-----AGCGCGCATTTAACGGTGAATC---CAAGAAAGCATGCT 4738
QY 75 SerProAlaTyrPheTyrValAlaAsnArgGlyAsnAsnGluGlyTyrAlaLeuValAla 94
DB 4739 AAAGTGGCGTCATTATTGTGCGC-----TTTCTCAATCTTGC 4777
QY 95 AlaAspAspArgIleProThrIleLeuAlaTyrSerProIleGlyArgPheAspMetasp 114
DB 4778 CCGTAAATGCGGTGGCGCATTTGGCTATCAATTAAGCTGGCGGACGTTATCGAATTGAC 4837
QY 115 SerMetProAspAsnLeuArgMetTyrLeuGlnIleTyrAspGlnGluIle-----Gly 132
DB 4838 AGC-----CGCACCTTTGATGAACGTGTGTTTCAAGGC 4870
QY 133 LeuIle-----LeuSerGlyLysAlaGlnLeuAsnGluIleLeuArgThrGlu 149
DB 4871 GTGTTCGCAATATGTTAACTAATCACTCAGCTGAATTCAGACCTGTTTATACAGT 4930
QY 150 GlyAlaProAlaGluValAlaAlaLeuMetAspAsnGlyHisPheAlaAsnAspPmet 169
DB 4931 CATATATGTCAGGCGCTTT-----GTTTGGTTTAAATACGCGCAT 4975
QY 170 ---ArgTTrpAsnGlnGlyTyrProTTrpAsnAsnLysGlu---ProLeuLeuProAsnGly 187
DB 4976 GGGCGGCTTTCTGTGATGCACCTTGTGCGACGCGTAATTCGCGTAAACATGTGAGC 5035
QY 188 AsnHisAlaTyr-----ThrGlyCysVal 195

DB 5036 AAAAGCGGTACAGCTTCGACGCGCACTTATAGTATTACTTAATGAAGTGGCACCAAT 5095
QY 196 AlaThrAlaAlaAlaGlnIleMetArgTyrHisSerTyrProLeu----- 210
DB 5096 ATCAGCTGGCAGCC-----TATCGCTATCTTCACGGGATTTTACACCTTAAGCAGC 5149
QY 210 ----- 210
DB 5150 ACCATTGGCTTACCGCACTTCACACAACTTAAACGGTGGCTATTTCCGTAAATTTAC 5209
QY 211 GlnGlyGluGlySerPheAspTyrHisAlaGlySerLeuValGlyAsnTyrSerGlyThr 230
DB 5210 CCGCCAAAATCACTTCAAGTAGTAAAGCCAAATCTGGGAATGGGAAATCTC 5269
QY 231 Phe-----GlyLumetCtyrAspTyrPheAsnMetProGlyAsnProAspLeuAspAsn 248
DB 5270 TATCTTTACGACCAACCTATTAATTTATTTGGGAAAAACGTCGACAGAAAT----- 5317
QY 249 LeuThrGlnSerGlnValAspAlaTyrAlaThrLeuMetArgAspValSerAlaSerVal 268
DB 5318 ---ACGCAATATCAAGTT---GCCATTCAAACAGCTCCACATTTCTTAATTAAGTCTGA 5371
QY 269 SerMetSer-----PheTyrGluAsnGlySerGlyThrTyrSerValTyrValAla 285
DB 5372 AACCTCTACAGAGATTATTAAGAAACGGGCAACGTCGACACACACATTTATTAAGT 5431
QY 286 GlyAlaLeu-----ArgAsnAsnPheArgTyrLysArgSer----- 297
DB 5432 CTCAGCCTGCCATTAGGCGATTAACCATTCGACATAGTATTTCTCCAGGCGTAC 5491
QY 298 -----LeuGlnLeuHisValAlaArgAlaLeuTyrThrSerGlnGluTyrPheAspMet 314
DB 5492 GATTTTAACCAACAGCTGGCGTAAATGGCTCTTT-----GGTGAACGTCAATCATG 5545
QY 315 IleArgGlyGluLeuAlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIle 334
DB 5546 AGTATGATGATTAACGCTTCA-----CGCAATATCAAA----- 5578
QY 335 GlyHisAlaPheValCysAspGlyTyrAlaSer---AspGlyThrPheHisPheAsnTyr 353
DB 5579 -----GGCTATCGCAGTTATGACGCTTAATCTTCGCTAATCAAT 5617
QY 354 GlyTyrPheGlyValSerAsnGlyPheTyrGlyLeuThrLeuLeuSerProHisSerLeu 373
DB 5618 AGCATTTGATGATAC---CGTCTCTTATTCACGATGATAGCTCTCAAAATCGCTCCATC 5674
QY 374 GlyIleGlyGlyGluGlyIle-----GlyPheThrIleTyrGlnGlu 387
DB 5675 TCACTGGCGCAAGCGGTGCTGTCGTGGCGCACAAACAGGATTAATACCTTAAGCCACT 5734
QY 388 IleIleThrGlyIleGluProAlaLysThrProAlaGluAlaGlyThrAspAlaLeuPro 407
DB 5735 GTT-----GGCAAAAGTTTTC 5752
QY 408 IleLeuAlaLeuLysAspIleGluAlaGluTyrLysSerGluSerGlyLeuAsnValGly 427
DB 5753 ATTATTCAGCGCAAAAGAT---GCCGAGAGAGAAAAGTGAATTCAGGTGCCAATGTGAGC 5809
QY 428 TyrSerIleTyrAsnThrGlyGluGlnGlnSerAsnLeuAspLeuGlyTyrArgLeuAsn 447
DB 5810 CTTGATTAATTC----- 5821
QY 448 LysAlaAspGlyGluValIleGluValLysThrSerSerIleAsnIleSerTyrGly 467
DB 5822 -----GGCAATGCGGTATGCTTACACACGCCGTATGAATTAATTAATGCGT 5872
QY 468 Tyr-----GlyGlnHisProGluSerPheSerLeuAlaProAsnGlnLeuSer 483
DB 5873 ATCAATTCATCATGATGCGAGGCGAATGTGAATTTGAAGCCATGAACGCCAAATCAT 5932
QY 484 GlnGlyIleAsnThrIleThrLeuLeuTyrArgArgThrGlyThrGlnGluGlnTyrPro 503

```



```

Db 5933 CCTCGTCAATTCATAGTATAGTATTCGCCAGCGCAAAAT----- 5980
Oy 504 VALATGHSALAGLNGLYGTYrValAsnSerLLeuValAsnThrAspProAsn 523
Db 5981 -----ACATATGGTGTATTAACTTCACTTGGCCAAAT 6013
Oy 524 AsnValVal-----ThrValAspAsnAsnGlyLysLeuSerLLeuVal 539
Db 6014 GCGGAGCCAGTCCCATGCGCATCCGCCAGCAAGAGG----- 6058
Oy 540 ProAsnSerPheValAlaAspLeuAsnSerTYrGluHisSerThrLethrValGlnPhe 559
Db 6059 -----GCATTGTGGCGCATGTG----- 6076
Oy 560 AsnSerAspSerProAspGluLLeuArgThrProValAlaPheAlaLeuSerThrGlyAla 579
Db 6077 -----GTGCAAGGTGGTGTGCTTTGCTTGAATAAATACCTACCCAG 6115
Oy 580 ThrAlaAspAspValLLeuSerLeuGlyTrp-ValMetAlaGluValProGlyGlySerSe 599
Db 6116 CCMAAAGCGAGTTATATC---GTCAATGGGTGGAGCGAGAAAGCAACATGCCGTTTC 6172
Oy 599 TAsnTYrProValValTrpSerLysAspValLeuThrLeuSerGluGlyAspTYrThrLe 619
Db 6173 CAATATCAAGTGAATTGG-----ATACGCA 6199
Oy 619 UTTPYrArgPheSerLLeuAsnGlnLysAspGluTrpLysLLeuGlySerValSe 639
Db 6200 CAAATACAAAGTCAGATATTCATACCAACCGCAAAATATATATATAGAGATTTA 6259
Oy 639 rVal-LysThrProThrGluTYrThrHisProLeuPheGluValGlyHisAsnGlnHis 659
Db 6260 TGCMAAAACACCCCAAAATTAATACC---GCCGTTTTC-----CATCAAAAT 6304
Oy 659 eTrhTYrThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLysAsnL 679
Db 6305 CCATGCTGCTGTTGAGT----- 6321
Oy 679 eugLYLeuProPheAsnGly-GluLeuValValPheArgGlnThr-GlnSerSerSe 698
Db 6322 --GGAGCAATATATAGTGAGCAAAATATATAGTGCCTTAAAGCTTAAAGTTCATCGTC 6379
Oy 698 rGlySerLeuTrpAlaAlaGlnGluThrValHisLLeuGlnGlyGluThrPheValY 718
Db 6380 TGGCTCTGCTGCTGCGT----- 6398
Oy 718 rLysProValValGluGlyProLLeuProAspGlySerTYrArgAlaThrLeuHisAlaPh 738
Db 6399 -----GGCTCT-----GCTTGAATGCGATTGT-- 6419
Oy 738 eValAsnGlyGlnGlnLeuTYrLeuLys-GlyLysArgAsnTYrThrValLysLLeuV 758
Db 6420 -----GGCACTGCTGCTTATGCTTACGATGCGAGTGCACCTTCAAGGGAGATTT 6472
Oy 758 al---AsnGlyThrAlaValAlaGluAlaLLeuSerSerGluGluLeuValPheProA 777
Db 6473 TAGAGTATGCGACTGT-----AAATGAAACACACGCCAAATTCGACAGGTTACCC 6526
Oy 777 snProAlaArgAspTYrValGluLLeuSer-----AlaProCysLLeuProGlnGlu 794
Db 6527 TCCCAACAGTGGAAAGCTAATTAAGCCACGACGCAACCGCCGCGTGCCT 6586
Oy 794 hSerLLeuLeuPheAspLeuSerGlyLysLLeuValMetLysAsnSerLeu-----S 812
Db 6587 TTTTCATCAGCTTAAAGATCAATGCAATGATGCTATGAAGCTAATCTGCTATTTA 6646
Oy 812 eAlaGlyHisGlyArgMetAspValSerArgLeuProAsn----- 825
Db 6647 AAGGGGAGACACACACAGGCAATCTTATCTTCCAAATTAAGCAGCAACGCCAAG 6706
Oy 826 -----GlyAlaTYrLLeuLysValAspGlyTYrThrThrLysLLeuAsnL 841
Db 6707 CCACCAACGTGGGCAATTCAAATTTGCAAGCGAGTGCATAGCAGCGCTATCAAGGTG 6765

```

```

RESULT 13
US-08-473-750-4
; Sequence 4, Application US/08473750
; Patent No. 5834187
; Patent No. 5834187
; Patent No. 5834187 5786143
; GENERAL INFORMATION:
; APPLICANT: Green, Bruce A.
; APPLICANT: Brinton, Jr., Charles C.
; TITLE OF INVENTION: Sequence and Analysis of LKP P111n
; Patent No. 5834187
; Patent No. 5834187 5786143
; TITLE OF INVENTION: Structural Gene and the LKP P111 Operon of No. 5834187 5786
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millita Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/473,750
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,321
; FILING DATE: 19-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Alice O.
; REGISTRATION NUMBER: 33,542
; REFERENCE/DOCKET NUMBER: ACC94-02B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9432 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (1882..2532)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2854..3630
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4016..6238
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6259..6873
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6955..8265
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 8395..9342
; US-08-473-750-4
;
; Alignment Scores:
; Pred. No.: 0.00115
; Score: 140.00
; Percent Similarity: 33.15%
; Best Local Similarity: 21.24%
; Length: 9432
; Matches: 196
; Conservative: 110
; Mismatches: 326

```


Db 6260 TGCACAAAACACCCAAAATTAACGCGCTTTTC-----CATCAAAAT 6304
Qy 659 eRThYrThrleuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLysAsnL 679
Db 6305 CCACCTCTACTGTACT----- 6321
Qy 679 eugLYleuProPheAsnGLY-GluLeuValValPheArgGlnThr-GlnSerSer 688
Db 6322 --GGACCAATTAATAGTGGCAATTAATAGTGGCTCAAAATGCTTAGGTTTCATCCTC 6379
Qy 698 rGLYSerLeuTrpAlaAlaGlnGlnThrValHisIleLysGlnGlyGlnThrPheValTy 718
Db 6380 TGGCTCTGCTGCTTCCGT----- 6398
Qy 718 rlySProValValGlnGlyProIleProAspGlySerTyArgAlaThrLeuHisAlaPh 738
Db 6399 -----GGCTCT-----GCTTGATTCATTCGT-- 6419
Qy 738 eValAsnGLYGlnGlnGlnLeuTyThrLys-GlyLysArgAsnTyThrValLysIleV 758
Db 6420 -----GGCACTGCTGCTTAGCTTAGCTTAGCTAGGAGAGTCAAGGAGAGATT 6472
Qy 758 aL---AsnGLYThrAlaValAlaGlnLysSerSerGlnGlnLeuArgValPhePro 777
Db 6473 TAGATGATGCACTGT-----AAATGAAACACACAGCCAAATCGACGTTACCC 6526
Qy 777 snProAlaArgAspTyValGlnLysSer-----AlaProCysIleProGlnGln 794
Db 6527 TGCACACAGTGGGAAAGCTTAATTTAAGCCACGACGCAAGCCGCCGCTGCTCCT 6586
Qy 794 hrSerIleLeuPheAspLeuSerGlyLysIleValMetLysAsnSerLeu-----S 812
Db 6587 TTTCATTCACGTTAAAGATGCAATGATGATGATGATGATGATGATGATGATGAT 6646
Qy 812 eValAGLYHisGLYArgMetAspValSerArgLeuProAsn----- 825
Db 6647 AAGGGGAGACACACACACAGGCAATCTTATCTTCCATATAGGACAGGCAAGGCAAG 6706
Qy 826 -----GlyAlaTyThrLeuLysValAspGlyTyThrLysIleAsnIle 841
Db 6707 CCACCAACGTGGCATTCAATTTGCAAGCCGATGCGATGCGATGCGATGCGATG 6765

RESULT 14
US-08-477-326-4
; Sequence 4, Application US/08477326
; Patent No. 5968769
GENERAL INFORMATION:
APPLICANT: Green, Bruce A.
APPLICANT: Branton, Jr., Charles C.
TITLE OF INVENTION: Sequence and Analysis of LKP P111
Patent No. 5968769
TITLE OF INVENTION: Structural Gene and the LKP P111 Operon of No. 5968769lypable
TITLE OF INVENTION: Haemophilus Influenzae
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: US
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,326
FILING DATE: 07-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/277,231

FILING DATE: July 19, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.
REGISTRATION NUMBER: 33,542
REFERENCE/DOCKET NUMBER: ACC94-02A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-9540
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 9432 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: complement (1882..2532)
FEATURE:
NAME/KEY: CDS
LOCATION: 2854..3630
FEATURE:
NAME/KEY: CDS
LOCATION: 4016..6238
FEATURE:
NAME/KEY: CDS
LOCATION: 6259..6873
FEATURE:
NAME/KEY: CDS
LOCATION: 6955..8265
FEATURE:
NAME/KEY: CDS
LOCATION: 8395..9342
US-08-477-326-4
Alignment Scores:
Pred. No.: 0.00115 Length: 9432
Score: 140.00 Matches: 196
Percent Similarity: 33.15% Conserved: 110
Best Local Similarity: 21.24% Mismatches: 326
Query Match: 3.15% Indels: 291
Gaps: 46
US-10-030-330-1 (1-843) x US-08-477-326-4 (1-9432)
Qy 22 AlaProValThrLysGlnArgAlaLeuSerLeuAlaArgLeuAlaLeuArg----- 38
Db 4571 GCCCAGTGGTGGCGTGGCGTGGCAACACAAAGCCAAAGTCAGATCAAAATGCGC 4630
Qy 39 -----GlnValSerLeuArgMetGlyGlnThrAlaValSerAspLysIleSer 54
Db 4631 TATACGATTTATTAATCAACACCGTCCCGCAGGCGCTTTCGTGATTAAGATTGTATG 4690
Qy 55 IleAspTyValTyArgGlnGlyAspAlaGlnArgGlyLeuThrSerGlnGlnGly 74
Db 4691 AGCGGTAT-----ACGGCGATTTAAGGTGAATC-----CAAGAAAGTGAAGT 4738
Qy 75 SerProAlaTyRheTyValAlaAsnArgGlyAsnAsnGlnGlyTyValAlaLeuValAla 94
Db 4739 AAGTGGCGTCAATTTATTTGCGCG-----TTTCTAATCTTGCC 4777
Qy 95 AlaAspArgGlnLeuProThrIleLeuAlaTyRSerProIleGlyArgPheAspMetAsp 114
Db 4778 CCCTTAATGCGTGTGGGCAATTCCTTATCAATTAACGCGATTCATTCGAATTCGAC 4837
Qy 115 SerMetProAspAsnLeuArgMetTrpLeuGlnIleTyArgGlnGlnIle-----Gly 132
Db 4838 AGC-----CGCACCTTATATGATACCTGCTTACAAAGC 4870
Qy 133 LeuIle-----LeuSerGlyLysAlaGlnLeuAsnGlnGlnIleLeuArgThrGln 149
Db 4871 GTGTTCGAATATGTTTAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 4930

OY 150 GlyValProAlaGluValAlaMetAspAsnGlyHisPheAlaAsnAspProMet 169
 DB 4931 CATTAATCGTCAGGCGCTTTT-----GGTTTGGTTTAAATACGGCAT 4975
 OY 170 ArgTPAsnGlnGlyTyrProTPAsnAsnLysGly---ProLeuPheProAsnGly 187
 DB 4976 GGGGCGTTTTCCTGATGTCACCTGGTCGACGCGTAATTCGCTAAACATGTGAC 5035
 OY 188 AsnHisAlaTyr-----ThyGlyCysVal 195
 DB 5036 AAAAAGCGCTACAGCTGCAGCGCATTAATGATTAACCTCAATGAAGAAGTGGCACAAT 5095
 OY 196 AlaThrAlaAlaAlaGlnLeuMetArgTyrHisSerTyrProLeu----- 210
 DB 5096 ATCAGTGGCAGCC-----TATCGTATTCCTTACGGGATTTTACACCTTAAGCGAC 5149
 OY 210 ----- 210
 DB 5150 ACCATTGGCTTAACCGCATTTTACAGACAATTTAGCGGTGGTATTTGCTGAATTTAC 5209
 OY 211 GlnGlyGlnGlySerPheAspTyrHisAlaGlySerLeuValGlyAsnTyrSerGlyThr 230
 DB 5210 CGCCCAAAAATCAAGTTTCAAGTGAAGTTTAAGCCAAAGCTGGGGAATGGGAAATCTC 5269
 OY 231 Phe-----GlyGluMetTyrAspTyrPheAsnMetProGlyAsnProAspLeuAspAsn 248
 DB 5270 TATCTTCAGACAAACACCTTAATTAATTTGGAAAAACGTGGCAGCAAT----- 5317
 OY 249 LeuThrGlnSerGlnValAspAlaTyrAlaThrLeuMetArgAspValSerAlaSerVal 268
 DB 5318 ---ACGCAATATCAGATT---GCCATTCAACAGCTGCACATTTCTTAATTACTGTGA 5371
 OY 269 SerMetSer-----PheTyrGluAsnGlySerGlyThyTyrSerValTyrValVal 285
 DB 5372 AACCTCACAAGATATGATTAAGAAACGGGCAACGTGACAAACGCAATTAATTAAGT 5431
 OY 286 GlyAlaLeu-----ArgAsnAsnPheArgTyrGlyAsnTyrSer----- 297
 DB 5432 CTCAGCTGGCATTAAGCGCATTAACATCTGCGAGATAGATTAATCTGCGAGGTGAC 5491
 OY 298 ---LeuGlnLeuHisValArgAlaLeuTyrThrSerGlnGlyTyrHisAspMet 314
 DB 5492 GATATTACCAACAGACTTGGCGTAATGGCTCTTTT---GGTGAACGTGATCAATGG 5545
 OY 315 IleArgGlyGluLeuAlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGlnSerIle 334
 DB 5546 AGTTAAGTATTACGCTTCA-----CGCAATTAATCAA----- 5578
 OY 335 GlyHisAlaPheValCysAspGlyTyrAlaSer---AspGlyThrPheHisPheAsnTyr 353
 DB 5579 -----GGCTATCGCAATTAATGACGTAATCTTTCGCAATACAAAT 5617
 OY 354 GlyTyrGlyGlyValSerAsnGlyPheTyrGlyLeuThrLeuLeuSerProThrSerLeu 373
 DB 5618 ACCATTGGTAGTTAC---CGTGGCTCTTATTCACGTGATACCTCAAAAATGGCTCATC 5674
 OY 374 GlyIleGlyGlyGlyIle-----GlyPheThrIleTyrGlnGlu 387
 DB 5675 TCACGGGCGCAAGCGGTGCTGTGCGGCAACACAGCTATTACCTTAAGCGAACAT 5734
 OY 388 IleIleThrGlyIleGluProAlaLeuThrProAlaGlnAlaGlyThrAspAlaLeuPro 407
 DB 5735 GTT-----GGCGAAAGTTTGGC 5752
 OY 408 IleLeuAlaLeuAspIleGlnAlaGlyTyrLysSerGlnSerGlyLeuAsnValGly 427
 DB 5753 ATTATTACGCCCAAAAGAT---GCCGCGAGGACAAAAGTGAATCAGGTGCACATGTGAC 5809
 OY 428 TyrSerIleTyrAsnThrGlyGlnGluGlnSerAsnLeuAspLeuGlyTyrArgLeuAsn 447
 DB 5810 CTGTGATTATTCT----- 5821
 OY 448 LysAlaAspGlyValIleGluValLysThrSerSerIleAsnIleSerTyrGly 467

DB 5822 -----GGCAATGGCGTTATGCTTAACACCGACCGCATGAATCAATTAATACGT 5872
 OY 468 Tyr-----GlyGlnHisProGluSerPheSerLeuAlaPheProAsnGlnLeuSer 483
 DB 5873 ATCAATTCATCTGATGGGAGGCGGATGTGGAAATTTGAAGCCATGAAACGCCAAATTCAT 5932
 OY 484 GlnGlyIleAsnThrIleThrLeuLeuTyrArgArgThrGlyThrGlnGlnIlePro 503
 DB 5933 CCTGCGCAATTAATCAATTAATGCTTAGTGAATTTCCGACGGCGGCAAAAT----- 5980
 OY 504 ValArgHisAlaGlnGlyGlyTyrValAsnSerIleLysValAsnThrThrAspProAsn 523
 DB 5981 -----ACATGGGTATTATTAACTCCTTGCACAAAT 6013
 OY 524 AsnValValVal-----ThyValAspAsnGlnGlyLysLeuSerIleVal 539
 DB 6014 GCGGACCGCATGCGCATGCGCATCCACCGCAACAGTACGAAAGG----- 6058
 OY 540 ProAsnSerPheValAlaAspLeuAsnSerTyrGlnHisSerThrIleThrValGlnPhe 559
 DB 6059 ---GCATTGTGGCGGATGTG----- 6076
 OY 560 AsnSerAspSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrGlyAla 579
 DB 6077 -----GTGCAAGGTGGTGTGCTTTGCTTAATTAACCTTAACCTTACCAG 6115
 OY 580 ThrAlaAspAspValIleSerLeuGlyTyrValMetAlaGluValProGlyLysSer 599
 DB 6116 CCAAAAGCGCATTAATC---GTCAAAATGGGTGAGGAGAAAGCGCAACATCCGTTTC 6172
 OY 599 AsnTyrProValValIleThrSerLysAspValLeuThrLeuSerGlnGlyAspTyrThr 619
 DB 6173 CAATATCAAGTGTGATTGG-----ATAACGCA 6199
 OY 619 TyrPyrArgPheSerIleAsnGlnLysAspGluTyrGlyLysIleGlySerValSer 639
 DB 620 CAATATCAAGTCAATGATTAATCAATGCAAAACGCAAAATTAATTAATTAAGGATTA 6259
 OY 639 Val-LysThrProThrGlyTyrThrHisPheLeuPheGluValGlnHisAsnGlnThr 659
 DB 6360 TGCAAAACCAACCCCAAAAATTAAC---CGCGTTTTC-----CATCAAAAT 6304
 OY 659 ThrTyrThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuLysAsn 679
 DB 6305 CCACTGCTACTGTACT----- 6321
 OY 679 LeuLysLeuProPheAsnGly-GluLeuValValPheArgGlnThr-GlnSerSer 698
 DB 6322 ---GGACCAATTAATTAATGAGCAAAATTAATTAATTAATTAATTAATTAATTAAT 6379
 OY 698 rGlySerLeuThrAlaAlaGlnGlnThrValHisIleLysGlnGlyLysThrPheValTyr 718
 DB 6380 TGGCTGCTGCTGCTGCTGCT----- 6398
 OY 718 LysProValValGlnGlyProIleProAspGlySerTyrArgAlaThrLeuHisAlaPhe 738
 DB 6399 -----GGCTCT-----GCTGATTCGCAATGT-- 6419
 OY 738 eValAsnGlyGlnGlnGlnLeuTyrLeuLys-GlyLysArgAsnTyrThrValLysIleVal 758
 DB 6420 ---GGACCTGCTGCTTATGCTTAAGTGGCAGAGTGCACCTTCAAGGAGGAGATT 6472
 OY 758 al---AsnGlyThrAlaValGlnAlaIleGluSerSerGlnGlnIleArgValPhePro 777
 DB 6473 TAAAGTAGGACACTGT-----AAATTAACAAAGACACCAAAATTCGACGGTAAACC 6526
 OY 777 snProAlaArgAspTyrValGlnIleSer-----AlaProCysIleProGlnGlnIle 794
 DB 6527 TGCACACAGTGGGAAAAGTAAATTAATTAAGCCACGAGGCAAAACCGCGCGCTGTGCTT 6586
 OY 794 hrSerIleIleLeuPheAspLeuSerGlyLysIleValMetLysAsnSerLeu-----S 812

OY	271	SerPheTYrGIuaNGISerGIYThrIYrSerValIYrValIaGLValaLeuKrrgSn	290
Db	851	AATTTCGCGGTCGGGGTAGT---TACAACGCTAAATCTCAGCATA-----	898
OY	291	AsnPheArGIYrLysArGIserLeuGln-----	299
Db	899	-----AGAACTCTTATCCGGGTGATATCACTACGGAAGATGAACCTGCCTCC	943
OY	300	-----LeuHisValaIArgAlaLeuTYrThrSerGlnGIuTrP-----	311
Db	944	TTGATTTATGTGGCTAGACGTAAACGATATATAT---TCATGCTCATGGGGTCCCGCTGAT	1000
OY	312	-----HispmetIleArvgGIuLeuAlaSerGIYArG	323
Db	1001	GACGGAGACATTTTACAAGCCCTTAGTACGTCGGGAAAAAGCGTTTAgTAAAAAGCT---	1057
OY	324	ProValTYrTYrAlaGIYAsnAsnGlnSerIleGIYHisAlaPheValaCysAspGIYTr	343
Db	1058	-----GTTACrGAGGAAGAAAGATTCCAAAGAGAGCATTTAGCT-----TTT	1099
OY	344	AlaSer-----AspGIYThrPheHisPheAsnTrpGIYTrpGIYValSerAnsGIY	361
Db	1100	GCCAGTGAATAATGGTGAACCTGCTGTATATATTCAAATTATACGAGCGCTATACTAATTC	1159
OY	362	PheTYrLysLeuThr-----LeuLeuSerProThrSerLeu	373
Db	1160	ATATATCTTATCTACTATTGGGGCTATTGATCACAAGATCTTACATCTCTTATTCGGA	1219
OY	374	GIYIleGIY-----GlyGIuGIYIleGIYPheThrIleTYrGln	386
Db	1220	GGTGTTCGCCGCGTCATGGCAGTCACGTATCTTCAAGTTCAGCGGAATATTTTATTCG	1277
OY	387	GIuIleIleThrGIYIle-----GIuProAlaLysThrProAla	399
Db	1280	AGTATATCAACGGCAGATGCAGTAATAGCCACGGTGAACGTCTCGCGCTCCCATTA	1339
OY	400	GIuAlaGIYThr-----AspAlaLeuProIleLeuAlaLeuLysAspIle	414
Db	1340	GCTGCGGGTGTTCACCTTGTCTTACTTAAGCCAAACCAACCTACTTGGAGAACGTGA	1399
OY	415	GIuAlaGIuTYrLysSerGIuserGIYLeuAsnValGIYTYrSerIleTYrAsnThrGIY	434
Db	1400	CAGTAT-----TTATCAATCTGTGCTCGGTA	1422
OY	435	GIuGIuGlnSerAsnLeuAsnLeuGIYTrArGIeAsnLysAlaAspGIYGIuValIle	454
Db	1427	GGGTTAGAAAGAACGCTGACGAGGAGATTTGGAGA---GATAGCGCATGGGGGANAAGAAATAC	1488
OY	455	GIuValIysThrSerSerIleAsnIleSerTrpGIYrGIYrGIYGIuHisProGIuser	474
Db	1484	TCTCATCGC-----TATGGCTTTGGTAAATCATGATCCCAT	1519
OY	475	PheSerLeuAlaProAsnGlnLeuSerGlnGIYIleAsnThrIleThrLeuLeuTYrArG	494
Db	1520	AAGTTAATTTGAATGTCCAAACACCTCGGAGATGTAAACGCAAAACCTGGTTTAACTCG	1577
OY	495	ArGIThrGIYThrGIuGlnTrpGIuProValaIArgHisAlaGlnGIYGIYrValaLysSer	514
Db	1580	CCAACA-----TTGATGTTTCCGAG	1600
OY	515	IleLysValaSnThrThrAspProAsnAsnValaValaIleThrValaAspAsnAsnGIuGIY	534
Db	1601	TCCACAAACTCCAGGAGAGACATATTAGATCCGTCATTAACATAT-----TCAAAAAA	1654
OY	535	LysLeuSerIleValProAsnSerPheValaAlaAspLeuAsnSerTYrGIuHisSerThr	554
Db	1655	AGTCTTCAA-----GATGCTAATCTCAAGACAAATTGAGCAGCTCACG	1699
OY	555	IleThrTYrAlaGlnPheAsnSerAspSerProAspGIuIleArGIThrProValaIaPheAla	574
Db	1697	GTACTGTGA-----GATATGTATACAGAAATTAGG-----	1722
OY	575	LeuSerThrGIYAlaThrAlaAspAspValIleSerLeuGIYTrpValaIaLeuAlaGIuVal	594

Dd	1727	-----GGAACTAGACACTGTGATTTAATATCA-----	1753
Oy	595	ProGlygly----SerSerAsnTyrProValTyrSerIlysAspValLeuThrLeuSer	613
Dd	1754	CCAGGGGGGAAATATTTCAAACCTTGGCGGTGTAAAGAACCAAGATGTT-----TCATCA	1807
Oy	614	GlucIysAspTyrThrLeuTyrThrArgPheSerIle-----AsnAsnGlnIlys	629
		: :	
Dd	1808	GAGGGA---TTCAAAGAAGCTGCACATTCATCTCTGTACACATTTGGGGTGAAGAGCGCGTA	1864
Oy	630	AspGluTyrPheIysIleGlySerValSerAlaIysThrProThrGluTyrThrHis---	648
Dd	1865	GGTATTTGGA-----ATCAGGTTAAGACA---ACAGAAATGACACAGG	1909
Oy	649	-----ProLeuPhe-----GluValGlnIHisAsnGlnThr	658
Dd	1910	ATTGACTTCCACACTTTGGAGGCTGAAGCTCTTTGGGAATCCATTCATTCATTCATTAACA	1969
Oy	659	SerThrTyrThrLeuAspMetAlaHisAsnArgValLeuProAspPheThrLeuIlysAsn	678
Dd	1970	GAACTTTTCGTCCTTTGGAAACGATAAAGAGAGGTTGAAACCAAGCTGTCAAGAAAGTACC	2029
Oy	679	LeuGlyLeuProPheAsnGlyGluLeuValValPheArgGlnThrGlnSerSerSer	698
Dd	2030	GTATCCACAAATTTGTGCCAGTTCAACTTCTATTTCATCCAGCGCTACTCTCAATCTTCT	2089
Oy	699	GlySerLeuTyrPheAlaIleGlnGluThrValHisIleGlyGlnGlyGluThrPheValTyr	718
Dd	2090	ATTCACAT-----GGGTGGAAACGTCGGCCATTCGCCCAAAGACACTACGCAATACC	2133
Oy	719	LysProValValGluGlyProIleProAsp-----GlySerTyrArgAlaThrLeuHis	736
Dd	2144	GATCCTGATGTCGTGATCCAAACACATCTCTTAAACCTTCTCTCTGAGCAAGCATGCAT	2203
Oy	737	AlaPheValAsn-----GlyGlnGlnGlnLeu-----Tyr	746
Dd	2204	TATTTTGTAAACAATATTTTGTGATTTGGCGCCACATTTTGTGTATATCTTCATCTTTT	2263
Oy	747	LeuIysGlyIysArgAsnTyrThrValIysIleValAsnGlyThrAlaValGluAlaIle	766
Dd	2264	ATGAAATCAAGAGAGAG-----ATCAGAAAGTCAAGAGCGGAAACGAT	2308
Oy	767	Glu-----SerSerGlnGluIleArgValPheProAsnProAlaArg	780
Dd	2309	GAATTCGATCATTCATATACAGACTCTGAGTACGATTCCTACTTTGGACAATGAACTTCC	2368
Oy	781	AspTyrValGluIleSerAlaProCysIleProGlnGluThrSerIleIleLeuPheAsp	800
Dd	2369	GGAATTAAGTGG-----CCCAAGAGGTTGAGAGACTTGCATTTGAT	2410
Oy	801	LeuSerGlyIysIleValMetIysAsnSerLeuSerAlaGlnIysGly	816
Dd	2411	TTTGCCGATGAAGACATCTTGCAAGTTGTCCTTCATCAAGAAACGGT	2458

RESULT 17
 US-09-453-702B-18
 : Sequence 18, Application US/09453702B
 : Patent No. 6365723
 :
 : GENERAL INFORMATION:
 : APPLICANT: Blatner, Frederick R.
 : Burland, Valerie
 : Perna, Nicole T.
 : Plunkett, Guy
 : Welch, Rod
 :
 : TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
 :
 : NUMBER OF SEQUENCES: 265
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Quarles & Brady
 : STREET: 1 South Plinckney Street
 : CITY: Madison
 : STATE: WI
 :
 : COUNTRY: US

```

      ZIP: 53701-2113
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Word Perfect 8.0
      CURRENT APPLICATION NUMBER:
      APPLICATION NUMBER: US/09/453,702B
      FILING DATE: 03-Dec-1999
      CLASSIFICATION: <Unknown>
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 60/110,955
      FILING DATE: 04-Dec-1998
      ATTORNEY/AGENT INFORMATION:
      NAME: Seay, Nicholas J.
      REGISTRATION NUMBER: 27386
      REFERENCE/DOCKET NUMBER: 960296.95017
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (608) 251-5000
      TELEFAX: (608) 251-9166
      INFORMATION FOR SEQ ID NO: 18:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 3695
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: linear
      MOLECULE TYPE: DNA (genomic)
      SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-453-702B-18

Alignment Scores:
      Pred. No.: 0.000279      Length: 3695
      Score: 139.50      Matches: 211
      Percent Similarity: 31.32%      Conservative: 111
      Best Local Similarity: 20.53%      Mismatches: 372
      Query Match: 3.14%      Indels: 336
      DB: 4      Gaps: 47

US-10-030-330-1 (1-843) x US-09-453-702B-18 (1-3695)
QY      63 AspAlaGluArgGlyLleThrSerGlnGluGlySerProAlaTyrPheTyrValAla 82
DB      |||.....|||.....|||.....|||.....|||.....|||.....|||.....
DB      526 GACGGTCACAGCCGCTTAATGCTTGTCTAACGCTAATGACCAAAACCGCTGCTGTG 585
QY      83 AsnATGGLyAsnAsnGlnGlyTyrAlaLeuValAlaAlaAspAspArgIleProThrIle 102
DB      ::::.....:::.....:::.....:::.....:::.....:::.....:::.....
DB      586 TCTCTGCGGAGCGGAGGCGCAGCATGACAGGCGCATGAAGATCAATCAAGACTGAA 645
QY      103 LeuAlaTyrSerProIleGlyArgPhe-----AspMetAspSerMetProAspAsn 119
DB      |||.....|||.....|||.....|||.....|||.....|||.....|||.....
DB      646 CTAACTTTCAAACCGGCTGGAATATGTGACTCGTCTCCCTGAAGGCCACATAATACACAG 705
QY      120 LeuArgMetTyrPleu---GlnIleTyrAspGlnGluIleGlyLeu-----133
DB      ::::.....|||.....:::.....|||.....|||.....|||.....|||.....
DB      706 GCAAAAGCAACACACTGGGTGAGTTACCGCAAACTGAAGCAGGGGTATCACTGTCTTT 765
QY      134 -----IleLeuSerGlyLyAlaGlnLeuAsnGlnGluIle---LeuArgThrGlu 149
DB      ::::.....|||.....:::.....|||.....|||.....|||.....|||.....
DB      766 ACTACCGGAAACGACAGTCAAGTCAAGCAACATTAAGTTCAGTTGATGACAGCAAA 825
QY      150 GlyValProAlaGluValAlaHisAlaLeuMetAspAsnGlyHisPheAlaAsnAspProMet 169
DB      |||.....|||.....|||.....|||.....|||.....|||.....|||.....
DB      826 ACCGTCACTGACAGACTGGGGCCAGCATGATGAT-----GTGGCAAACTCCACCCCTG 879
QY      170 ArgTyrAsnGlnGlyTyrProTyrAsnAsnLyGluPro-----LeuLeuProAsn 186
DB      |||.....|||.....|||.....|||.....|||.....|||.....|||.....
DB      880 AGCCCTAAC-----GAGCCGTCAAGTACAGCTGGTTGCTGAT 915
QY      187 GlyAsnHisAlaTyrThr-----GlyCysValAlaTat 197
DB      |||.....|||.....|||.....|||.....|||.....|||.....|||.....
DB      916 GGTTCAGCAACCTTAATACGTTGACGTTGACTGCGGTGACTCCAGAGGTAAATCCGGTGACG 975
QY      198 AlaAlaAlaGlnIleMetArgTyrHisSer-----Tyr---208

```

```

DB      976 GGAGAAAGCCAGCCGCTTGGCATTTGT-TCGGCAAGACACTAATGGTGAACCGTTGGTGC 1034
QY      209 -----ProLeuGlnGlnGlyLysSerPheAsp-----217
DB      1035 CATTTGGAAATTAACACAGCGGTTACAGCGCGGGGTTTCTTGACCCCGCCGGA 1094
QY      218 -----TyrHisAlaGlySerLeu-----223
DB      1095 CGTTTGTGGCTGCTTTCAGCGAGCATGATGACTGGGACATTACAAACAGCTGAA 1154
QY      224 -----Val 224
DB      1155 GTTTGTGGCGGGCGCTTGATGACACATTCGTCATCAACCCGTAATCTGATAACC 1214
QY      225 GlyAsnTyrSerGlyThrPheGlnGlyLeuMetTyrAspTyrPheAsnMetProGlyAsnPro 244
DB      |||.....|||.....|||.....|||.....|||.....|||.....|||.....
DB      1215 GGTGGTTGGGGGACAGTTACGCG-----AATCTGACGGTAAAGA 1256
QY      245 AspLeu-AspAsn---LeuThrGlnSerGlnValAspAlaTyrAlaThrLeuMetArgAs 263
DB      |||.....|||.....|||.....|||.....|||.....|||.....|||.....
DB      1257 TGCTTATGACAAACCTGTGACACAGCCTCAGCCCGGAAACCCGTCATTAGCGGTGCCG 1316
QY      263 PValSerAlaSerValSerMetSerPheTyrGluAsnGlySerGlyThrTyrSerValTyr 283
DB      |||.....|||.....|||.....|||.....|||.....|||.....|||.....
DB      1317 TGCTGAAGCTTCTACGCGATCGGCGCTGGACAATAATATGATGAGTGAAGTGAAGTGC 1376
QY      283 rVal-ValGlyAlaLeuValArgAsn---AsnPheArgTyrLysArgSer-----297
DB      ::::.....|||.....|||.....|||.....|||.....|||.....|||.....
DB      1377 GATTACTCTCGCTCTACGGCGGCTGAATTAAGATTATGCGCAAGCTTAATGACAGAA 1436
QY      298 -----LeuGlnLeuHis-----ValArgAlaLeuTyrThrSerG 309
DB      |||.....|||.....|||.....|||.....|||.....|||.....|||.....
DB      1437 TCGGCAAGCAATCGCGCAAAATGTAACCGTGTGGCTGATGCTTATCTCAACAGATC 1496
QY      309 GlnGlu-----311
DB      ::::.....
DB      1497 GAAAGTCTCTGTGCGCAGAAATCAGTAAACCGCGGAAAGCAACACCGTGAAGCTGTG 1556
QY      311 rPHisAspMetIleArgGlyGlnLeuAlaSerGlyArgProValTyrTyrAlaGlyAsnA 331
DB      |||.....|||.....|||.....|||.....|||.....|||.....|||.....
DB      1557 GCGCAAAAGATGCG-CATGACAGCGTATCATGCTGT-----1590
QY      331 snGlnSerIleGlyHisAlaPheValCysAspGlyTyrAlaSerAspGlyThrPheHisP 351
DB      ::::.....|||.....|||.....|||.....|||.....|||.....|||.....
DB      1591 -----CTTGGCTGTGCGCAAGTTTGAACGGGGACCCGCTTGAAGGGGAGCAGCGTTT 1642
QY      351 heAsnTyrPglTyrGlyGlyValSerAsnGlyPheTyrLysLeuThrLeuSerProT 371
DB      ::::.....|||.....|||.....|||.....|||.....|||.....|||.....
DB      1643 CCAAGTTGACCGAAAGATG---AACGGTTCATATGTGCTACGTTGACTACACAGTG 1696
QY      371 hrSerLeuGly-----374
DB      |||.....
DB      1697 GAAAGAGGGGAGCTTCCGCTCAAGCTCTCTTCAACGGCCAGCAGCAGCAGCAGGAAG 1756
QY      375 -----IleGlyGlyGlu-----378
DB      |||.....|||.....|||.....|||.....|||.....|||.....|||.....
DB      1757 CCGCGCAGTTGACGGTCAATCGCGGAGAGATCATCAAGCAACTTCAGCTTGTGGCG 1816
QY      379 -----GlyIleGlyPheThrIle-----T 385
DB      ::::.....|||.....|||.....|||.....|||.....|||.....|||.....
DB      1817 ACAATTAAGCTCCGACCGTCAAAACGACAGCAGACACTCACCTTGTGGCG 1876
QY      385 yrcGlnGluIleThrGlyIleGluPro-----AlaLysT 397
DB      |||.....|||.....|||.....|||.....|||.....|||.....|||.....
DB      1877 ACGGGAACCGGTACCCGGGTGAAGCCAGATGACCAAGTGTTAAGCGGTCCCGCCAGCA 1936
QY      397 hrProAlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGlnIag 417
DB      ::::.....|||.....|||.....|||.....|||.....|||.....|||.....
DB      1937 CGGGAGTGAAGCGTCTTCA-----GAG 1960
QY      417 LuTyrLysSerGlnSerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGlyGlnGluG 437
DB      ::::.....|||.....|||.....|||.....|||.....|||.....|||.....

```



```

Db 1961 GAAACTGACAGAGAAAGTAAATGCGGTCTACGTGACCTTAACGCTGGATCGCG 2020
QY 437 InsertionLeuaspLeuGlyTyrArgLeuasn----- 447
Db 2021 CCGGTGACGTGTCTGATGCCGCGAGTACAGCGCCAAAATGCCGTGCTACGCCAC 2080
QY 448 -----TysalaAspGlyGluValIleGluValIleThrSerIleAsnIleG 464
2081 TCGTGAACCTTCGACGTGACGCATTAAGCTGAGATTGCTGATTACAGTAAGCTTA 2140
QY 464 eTTPtGlyTyrGlyGluHisProGluSerPheSerLeuAlaProAsnGlnLeuSerG 484
2141 AT----- 1AACCACCTGGCTA 2155
Db 484 InGly-----IleAsnThrIleThrLeuLeuTyrArgArgThrGluGlnTrpG 502
2156 ATGACAGCTGCTGCTAACCAGATTAACCTG----- 2184
QY 502 IuproValArgHisAlaGlnGlyTyrValAsnSerIleLysValAsnThrThrAsp 522
2185 --ACCGTTGTGACACCTATGGTAACCCGTGACGGCGACGAGAGTTAGCCTGACTTAC 2242
522 roAsnAsnValVal-----ValThrValAspAsnGlnGlyLysLeus 537
2243 CCGAGGGTGTGACACGACGAGCGGGAATACAGTAACACTAATGCGCGCATTAAGCGG 2302
QY 537 eTlleValProAsnSerPheValAlaAspLeuAsnSerTyrGluHisSer----- 553
2303 ACATTGAGCTATGTTCAACGGTTCGCGGAGAACACAAATATTCCTCGCTGGTGAATG 2362
QY 554 -----ThrIleThrValGlnPheAsnSerAsp---SerProAspGluIleArgThrP 570
2363 CTCAGAGACGGTGCACGCGTAATTCACGCGGATCCACGACCGGTGCAGCAAACTGC 2422
QY 570 roValAlaPheAlaLeuSerThrGlyAlaThrAlaAspAspValIleSerLeuGlyT 590
2423 AGCTGAGACCGCGCTGCTCAAAAAGTGGCAACGCGCAAAAGATGCTTACGTCAGCG 2482
QY 590 alMetAlaGlu-----ValProGlyGly-----SerSerAspTyrPro- 602
2483 ACCTTGAGAGATAAAATGTTAACCTGTTCCAGGGAGCCTGTGACCTTAATCTGCC 2542
QY 603 -----ValValTyrSerLys----- 607
2543 GGGGTGTCAAGCCGTTACAGCGGATTAATGTCGGGTAAAGCCACATGAGGGGAAG 2602
QY 608 -----AspValLeuThrLeuSerGlnGlyAspTyrThrLeuTyrTyrArgPheSerI 625
2603 CAGAGTTGCAGGTGTTTTCAGTACTGCCGGAACGTATGATCAGCGCATCGCAGG- 2661
625 IeAsnAsnGlnLysAspGluTyrPylsLysIleGlySerValSerValLysThrProThG 645
2662 --AATAGCCAGCCTTCGAATACGACAGATTAACGTTTGTAGCCGTAAGGCTACCGCA 2719
QY 645 IuTyrThr-----HisProLeuPheGluValIleGlnHisAsnGlnI 658
2720 CCGTCTCCGATTTAGTGGATGATGGAATATGACCTGCGCGGCAATGCCAAACAGA 2779
QY 658 hSerThrTyrThrLeuAspMetAlaHisAsnArgValLeuProAsp-----PheThrL 676
2780 CCGTAAAGTATGACGGATGATGCCAATTAACAACTGTTGAAAGATAGAGAGTACCG 2839
QY 676 eu-----LysAsnLeuGlyLeuProPheAsnGlnGlyGluLeuValValPheA 692
2840 TGAATGCCAGCCCGGCAAAATTTAGTTCTACTGCCCAATGGAGC-----G 2884
QY 692 rgGlnThrGlnSerSerGlySerLeuTyrPalaIleGlnGluTyrValHisIleLysG 712
2885 CGAAACTAATGACAGACAGAGGCTATTTTACCGCCACGACCACTGTGCGCGCAAA- 2943
QY 712 InGlyGluThrPheValTyrLysProValValGlnGlyProIleProAspGly-----S 730
2944 -----TATACACTCAGCGCGAAAGTAGATCAGCGCGACGCTACAGAAAT 2986

```

```

QY 730 eTyrArgAlaThrLeuHisAlaPheValAsnGlnGlnGlnLeuTyrLeuLysGlyL 750
Db 2987 CGACGAAACTGCCGAATTAATTCGTGCGGATGATACAAATGCAGTACTACCGCAT 3046
QY 750 ysArgAsnTyrThrValLysIleValAsnGlyThrAlaValIleGluIleSerSerG 770
Db 3047 CATCTGATGATCTTCTCTGCTGCGGATGCGGATGCGATTCGACCTGGAGGTGACAC 3106
QY 770 InGluIleArgValIlePheProAsnProAlaArgAspTyrValIleGluIleSerAlaPro- 788
Db 3107 TGAATGCGCAATTAACCCGTTGGGGGAATATGTGCGTACATTAAGACCGCAAG 3166
QY 789 -----CysIleProGlnGlnThrSerIleIleLeuPheAspL 801
Db 3167 GGGTGAACGAGAAAGATTATCACTTCCTGCGCAAAATGACCATTTTC-----G 3217
QY 801 euserGlyLysIleValIleMetLysAsnSerLeuSer---AlaGlyHisGlyArgMetAspV 820
Db 3218 TGAGCGGAAATATCAGCGCGTACATTTAGTACCAAGCAAGCCTGGTGTCTATACGTCAT 3277
QY 820 alSerArgLeuProAsnGlyAlaTyrIleLeuLys-----V 832
Db 3278 TTAACCCCTGACGTATGCGCGGTACGAAATGAAGCCAGTGAAGCATTACCGCGG 3337
QY 832 alAspGlyTyrThrThrLys 838
Db 3338 TGGATCCGATACGCAAG 3357

RESULT 18
US-08-702-572-14
; Sequence 14, Application US/08702572
; Patent No. 5965386
; GENERAL INFORMATION:
; APPLICANT: Kerry-Williams, Sean M
; APPLICANT: Gilbert, Sarah C
; TITLE OF INVENTION: Yeast Strains and Modified Albumins
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Centeon L.L.C.
; STREET: 1020 First Avenue
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-1310
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Microsoft Word 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/702, 572
; FILING DATE: 11-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: NO 95/23857
; FILING DATE: 1-MAR-1995
; APPLICATION NUMBER: GB 9404270.2
; FILING DATE: 5-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Naomi Biswas
; REGISTRATION NUMBER: 38,384
; REFERENCE/DOCKET NUMBER: CE0114 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610/878/4294
; TELEFAX: 610/878/4221
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4106 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

```


Db 3303 GATCCGATTCGATCCAAACACCTCTTAATAAACTTTCTCCCTCCGTAAGGACCATGATC 3362
QY 737 AlaphaveAlasn-----GlyGingIngleu-----Tyr 746
Db 3363 TATTTTAAACATATTTTGTGATGGCCACATTTTGTGTATCTCATGTTT 3422
QY 747 LeuLysGlyLysArgSerThrValLysIleValAsnGlyThrAlaValAlaIle 766
Db 3423 ATGAAATCAAGGAAAGG-----ATCAGAGGTCAAGGCGGAAACGTAAT 3467
QY 767 Glu-----SerSerGluGluIleArgValPheProAsnProAlaArg 780
Db 3468 GAATTCGATATCATGATGATACAGACTCTGATGATGATTCATCTTGACATATGAACTTCC 3527
QY 781 AspTyrValGluIleSerAlaProCysIleProGluGluThrSerIleIleuPheAsp 800
Db 3528 GGAATTTACTGAG-----CCCGAAGAGGTGAGGACTTCGATTTTGAT 3569
QY 801 LeuSerGlyLysIleValMetLysAsnSerLeuSerAlaGlyHisGly 816
Db 3570 TTGTCGATGAGAAACACATCTTGCAAGTTGTTCTTCATCAGAAAACGCT 3617
RESULT 19
5521093-4
Patent No. 5521093
APPLICANT: LEMOINE, YVES; NGUYEN, MARTINE; ACHSTER, TILMAN
TITLE OF INVENTION: YEAST VECTOR CODING FOR HETEROLOGOUS
GENE FUSIONS LINKED VIA KEX2 CLEAVAGE SITE AND CODING FOR
TRUNCATED KEX2 GENES
NUMBER OF SEQUENCES: 8
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/393, 025
FILING DATE: 23-FEB-1995
Prior Application Number: 191, 354
FILING DATE: 07-FEB-1994
APPLICATION NUMBER: 26, 121
FILING DATE: 04-MAR-1993
APPLICATION NUMBER: 500, 885
FILING DATE: 29-MAR-1990
SEQ ID NO: 4
LENGTH: 4732
5521093-4
Alignment Scores:
Pred. No.: 0.000418 Length: 4732
Score: 139.50 Matches: 167
Percent Similarity: 32.67% Conservative: 80
Local Similarity: 22.09% Mismatches: 278
Match: 3.14% Indels: 231
Gaps: 43
US-10-030-330-1 (1-843) x 5521093-4 (1-4732)
QY 166 AsnAspPrometArgTyrPasnGlnGlyTyrProTyrPasnAsnLysGluProLeuPro 185
Db 1748 AATGATCCGCTTTTGAGAGGAG-----TGACACTTGCTCATCAATTTTCTT 1798
QY 186 Asn-----GlyAsnHisAlaTyrThrGlyCysVal 195
Db 1799 GGCAGATGATATAATGTTCTTGATCTGTGTACAAATATAATACAGCGCGAGGGCGT 1858
QY 196 AlaThrAlaAlaAlaGlnIleMetArgTyrHisSerTyrProLeuGln----- 211
Db 1859 GCGCCATTTGATGATGCGCTTGACTAGCAAAAGAAAGACCTTGAGGATTAATTTTGC 1918
QY 212 GlyGluGlySerPheAspTyrHisAlaGlySerLeuValGlyAsnTyrSerGlyThrPhe 231
Db 1919 GCGGAAGGTTCTTGATTTCAAC----- 1942
QY 232 GlyGluMetTyrAspTyrPheAsnMetProGlyAsnProAspLeuAspAsnLeuThrGln 251
Db 1943 -----GACAAATACCAATTTTACCT-----AAACCAAGATTTATCTGAT----- 1978

QY 252 SerGlnValAspAlaTyrAlaThr-----LeuMetArgAspValSerAlaSerValSerMet 270
Db 1979 -----GACATACATGCTGACAGATGTCAGGTGAATTAATCTGCAAAAAAGGTAC 2029
QY 271 SerPheTyrGluAsnGlySerGlyThrTyrSerValTyrValAlaGlyAlaLeuArgAsn 290
Db 2030 AATTTTGGCGGTCCGGGTAGGT-----TACACGCTAAATCTCAGGCATA----- 2077
QY 291 AsnPheArgTyrLysArgSerLeuGln----- 299
Db 2078 -----AGAACTTATCCCGGTATATCATCTACAGGAAGATGAAGCTGCCTCC 2122
QY 300 -----LeuHisValArgAlaAlaLeuTyrThrSerGlnIleTyr----- 311
Db 2123 TTGATTTATGTTCTAGACGTAAACGATATATAT-----TCAGTCAATGAGGCTCCCGCTGAT 2179
QY 312 -----HisAspMetIleArgGlyGluLeuAlaSerGlyArg 323
Db 2180 GAGGAAAGACATTTACAGGCCCTAGTACCTGCTGTAAGAAAGGCTTTAGTAAAGT----- 2236
QY 324 ProValTyrTyrAlaGlyAsnAsnGlnSerIleGlyHisAlaPheValCysAspGlyTyr 343
Db 2237 -----GTTACTGAGGAGAGATTCACAAAGAGCGATTTACGTT-----TTT 2278
QY 344 AlaSer-----AspGlyThrPheHisPheAsnTyrProGlyGlyValSerAsnGly 361
Db 2279 GCCAGTGGAAAGTGGTGAAGCTCGTGTATTAATTCAGACGGCTTATCAATATCC 2338
QY 362 PheTyrLysLeuThr-----LeuLeuSerProThrSerLeu 373
Db 2339 ATATATTCATATCTACTATTTGGCGCTATGATCACAAGATTCATCTCTTATTCGCA 2398
QY 374 GlyIleGly-----GlyGluGlyIleGlyPheThrIleTyrGln 386
Db 2399 GGTGTGCCCGCCCTCAGTCAGTCAGCTATTTCTTACAGTTCAGCGCAATATATCAATCC 2458
QY 387 GluIleIleThrGlyIle-----GluProAlaLysThrProAla 399
Db 2459 AGTGATATCAACGCGAGATGACAGTAAATAGCCACGCGTGAACGCTGCGCTCCATTA 2518
QY 400 GluAlaGlyThr-----AspAlaLeuProIleLeuAlaLeuLysAspIle 414
Db 2519 GCGTCCGCGTGTACACTTTGTTACTAGAACCAACCAACCACTTATGAGAGAGCTA 2578
QY 415 GluAlaGluTyrLysSerGlySerGlyLeuAsnValGlyTyrSerIleTyrAsnThrGly 434
Db 2579 CAGTAT-----TTATCAATCTGCTGCGGCTA 2605
QY 435 GluGluGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAlaAspGlyValIle 454
Db 2606 GGGTTAGAAAGAACGCTGACGAGATTTGAGA-----GATAGCGCATGGGAAAGAAATAC 2662
QY 455 GluValLysThrSerSerIleAsnIleSerTyrGlyTyrGlyGluHisProGluSer 474
Db 2663 TCTCATCGC-----TATGCGTTTGGTAAATCATGATGCCAT 2698
QY 475 PheSerLeuAlaProAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuTyrArg 494
Db 2699 AACTTAATTCGAATGTCACAGACCTGGGAAAGATTTACACACAAACCTGTTTACCTG 2758
QY 495 ArgThrGlyThrGluGlnTyrProValArgHisAlaGlnGlyGlyTyrValAsnSer 514
Db 2759 CCACACA-----TTGTATGTTTCCAG 2779
QY 515 IleLysValAsnThrThrAspProAsnAsnValValAlaThrValAspAsnAsnGly 534
Db 2780 TCCAAACATCCACAGAGAGACATTTGATTCGCTATACATA-----TCAGAAAA 2833
QY 535 LysLeuSerIleValProAsnSerPheValAlaAspLeuAsnSerTyrGluHisSerThr 554
Db 2834 AGCTTCA-----GATGCTAACTTCAAGAAATAGAGACGATCAGC 2875

```

OY 555 IIEthValGlnPheanSerProaspGluIleargThrProValAlaPheAla 574
   :|||||
Db 2876 GFAACTGTA-----GATATTGATACGAAATTAGG----- 2905
OY 575 LeuSerThrGlyAlaThrAlaAspValIleSerLeuGlyTrpValMetAlaGluVal 594
   :|||
Db 2906 -----GGAAGCTACGACTGCTGATTAAATATCA----- 2932
OY 595 ProGlyGly----SerSeranTrpProValIleTrpSerLysaspValLeuThrLeuSer 613
   :|||
Db 2933 CCAGCGGGGATATTTCMAACCTTGCGCTGTAGACCAAGACATGTT-----TATCA 2986
OY 614 GluGlyAspTrpThrLeuTrpTrpArgPheSerIle-----AsnAsnGlnLys 629
   :|||
Db 2987 GAGGGA---TTCMAAAGACTGACATTCATGTCTGTAGACATGTGGGTGAGAACGGCGTA 3043
OY 630 AspGluTrpLysLysIleGlySerValSerValLysThrProThrGluTrpThrHis--- 648
   :|||||
Db 3044 GGTGATTGGAAA-----ATCAAGGTTAAGACA---ACAGAAAATGGACACAGG 3088
OY 649 -----ProLeuPhe-----GluValGlnHisAsnGlnThr 658
Db 3089 ATTGACTTCACAGTTGGAGGCTGAAAGCTCTTTGGGAAATCCATTGATTCATCTTAACA 3148
OY 659 SerThrTrpThrLeuAspMetAlaHisAsnArgValIleuProAspPheThrLeuLysAsn 678
   :|||||
Db 3149 GAAACTTTCGCTCTTGGAAACGATAAAGAGAGGTTGAACACAGCTCTACAGAAATAC 3208
OY 679 LeuGlyLeuProPheAsnGlyLysLeuValValPheArgGlnThrGlnSerSer 698
   :|||
Db 3209 GATATCAAAATTTCTGCCAGTTCCAACTTATTTCCATGAGCGCTCTTCTCATATCTCT 3268
OY 699 GlySerLeuTrpAlaIleGlnGluThrValHisIleLysGlnGlyLysThrPheValTyr 718
   :|||||
Db 3269 ATCTCAAT-----GGTGTGAAAGCTGGCCATCCCAAAAGCACTACTGGAGTACC 3322
OY 719 LysProValValGluGlyProIleProAsp-----GlySerTrpArgAlaThrLeuHis 736
   :|||
Db 3323 GATCTGATTCGTATCCAAACACTCTTAATAAACTTCTCTCCATGCGAAGCAAGCAT 3382
OY 737 AlaPheValAsn-----GlyGlnGlnGlnLeu-----Tyr 746
   :|||||
Db 3383 TATTTTAAACAATATTTTGAATGGCCGACATTTTGGTGTATTAATCTCATGTTT 3442
OY 747 LeuLysGlyLysArgAsnTrpThrValLysIleValAsnGlyThrAlaValGluAlaIle 766
   :|||||
Db 3443 ATGAAATCAAGAGGAAG-----ATCAGAAGGTCAGAGCGGAAGCAT 3487
OY 767 Glu-----SerSerGluGluIleArgValPheProAsnProAlaArg 780
   :|||
Db 3488 GAATTCGATATCATGATGATACAGACCTCTGAGTACGATTCATCTTGGACAAATGCACTCC 3547
OY 781 AspTrpValGluIleSerAlaProCysIleProGlnGluThrSerIleLeuPheAsp 800
   :|||
Db 3548 GCAATTACTGAG-----CCGGAAGAGGTTGAGGACTTCGATTTTGAT 3589
OY 801 LeuSerGlyLysIleValMetLysAsnSerLeuSerAlaGlyHisGly 816
   :|||||
Db 3590 TTGTCCGATGAAGACCATCTTGCAAGTTGTCTCTCATCAGAAAACGCT 3637

```

```

RESULT 20
US-08-750-532-8
: Sequence 8, Application US/08750532
: Patent No. 5756339
GENERAL INFORMATION:
: APPLICANT: MITTA, Masanori
: APPLICANT: YAMAMOTO, Katsuhiko
: APPLICANT: MORISHITA, Mio
: APPLICANT: ASADA, Kiyozo
: APPLICANT: TSUNASAWA, Susumu
: APPLICANT: KATO, Ikunoshin
: TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENE
: NUMBER OF SEQUENCES: 18

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
STREET: 419 Seventh Street N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/750,532
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP95/01095
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 1994/130236
FILING DATE: 13-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 1994/173912
FILING DATE: 26-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: MITTA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ. ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 4765 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-750-532-8
Alignment Scores:
Pred. No.: 0.00609 Length: 4765
Score: 138.00 Matches: 163
Percent Similarity: 31.13% Conservative: 87
Best Local Similarity: 20.30% Mismatches: 264
Query Match: 3.11% Indels: 290
Gaps: 42
US-10-030-330-1 (1-843) x US-08-750-532-8 (1-4765)
OY 185 ProAsnGlyAsnHisAlaTrpThrGlyCysValAlaThrAlaAlaGlnIleMetArg 204
   :|||||
Db 1174 CCTACGGGAATATGACATATTTGG----- 1200
OY 205 TyrHisSerTrpProLeuGlnGlySerPheAspTrpHisAlaGlySerLeuVal 224
   :|||
Db 1201 -----TGGAGTGTACAGGTGACGGAAGCT-----CACGTGAGCTGGAAGTGTCT 1245
OY 225 Gly-----AsnTrp-----SerGly----- 229
   :|||
Db 1246 GGTTCAGACAGCAACATGATCTTGGGATTTGGCTCAGTATGACTGTGTAATGGGA 1305
OY 230 ThrPheGlyGluMetLysAspTrpIleAsnMetProGlyAsnProAspLeuAsnLeu 249
   :|||||
Db 1306 GTGTTCTCAAGACTATGCTTGG-----GATTATACGAACGTT 1344
OY 250 ThrGln-----SerGlnValAspAlaTrpAlaThrLeu 260
   :|||
Db 1345 ACCACAGACACCGTGACGGGTGTGCTCCAGGTGCCAAATATGCAATAGAGTCTT 1404
OY 261 MetArgAspValSerAlaSer----- 267

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: June 1, 2003, 22:34:57 ; Search time 2893 Seconds

(without alignments)
4719.252 Million cell updates/sec

Title: US-10-030-330-1

Perfect score: 4438

Sequence: 1 MKKSLFLAIVMLFGIAMQGH.....PNCAYILKVDGYTITINIVH 843

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Matched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus.model -DEV=xlh
-Q/cgn2.1/USPTO_spool/US10030330/runat_23052003_181645_6762/app_query.fasta.1.1031
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.ccl -LIST=45
-LOCAL=IGR-200 -THR=SCORE-DECT -THR_MAX=100 -THR_MIN=0 -ALIGN=20 -MODE=LOCAL
-OUTFMT=ptc -NOM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10030330@cgn2.1.1.2087 -runat_23052003_181645_6762 -NCPU=6 -ICPU=3
-NO_XLPTX -NO_MAP -LARGEQUERT -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEBOU=120 -WARN_TIMEBOU=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: EST:
2: em_estbta:
3: em_estbta:
4: em_estbta:
5: em_estbta:
6: em_estbta:
7: em_estbta:
8: em_estbta:
9: em_estbta:
10: em_estbta:
11: em_estbta:
12: em_estbta:
13: em_estbta:
14: em_estbta:
15: em_estbta:
16: em_estbta:
17: em_estbta:
18: em_estbta:
19: em_estbta:
20: em_estbta:
21: em_estbta:
22: em_estbta:
23: em_estbta:
24: em_estbta:
25: em_estbta:
26: em_estbta:
27: em_estbta:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	123.5	2.8	1667	11	AY103544
2	114	2.6	1557	17	BH770729
3	111	2.5	2178	11	AK018613
4	104	2.3	1093	13	BM468550
5	103.5	2.3	3170	11	AK004957
6	102.5	2.3	820	9	AU140457
7	102.5	2.3	1594	17	BH770545
8	101	2.3	912	12	BG245705
9	101	2.3	730	12	BG251484
10	100.5	2.3	2530	11	AK018522
11	100.5	2.3	2660	11	AK018132
12	99.5	2.2	765	17	BH559216
13	99	2.2	563	13	BI934852
14	99	2.2	894	17	BH158271
15	99	2.2	944	13	BI951857
16	98.5	2.2	765	14	BM815828
17	98.5	2.2	1378	12	BF383380
18	98	2.2	792	12	BG886970
19	98	2.2	1037	17	CNS077A6
20	98	2.2	1149	13	BI838400
21	97.5	2.2	624	13	BM590344
22	97.5	2.2	688	13	BM609603
23	97.5	2.2	707	13	BM583419
24	97.5	2.2	718	13	BM609445
25	97.5	2.2	727	17	AQ449954
26	97.5	2.2	913	17	CNS06M7S
27	97.5	2.2	1530	11	AY104834
28	97	2.2	652	17	PT025609R
29	96.5	2.2	499	12	BG455676
30	96.5	2.2	524	12	BG560530
31	96.5	2.2	1130	12	BG752175
32	96.5	2.2	1241	13	BM547472
33	95.5	2.2	1675	11	AY103967
34	95.5	2.2	2637	9	AK019752
35	95	2.1	610	9	AJ499355
36	95	2.1	884	17	AZ673465
37	95	2.1	928	14	BQ710481
38	95	2.1	2949	11	BC006017
39	94.5	2.1	581	12	BG301151
40	94.5	2.1	633	13	BM634464
41	94.5	2.1	701	10	AW584470
42	94.5	2.1	712	13	BM593829
43	94.5	2.1	746	12	BG548053
44	94.5	2.1	962	10	BE035942
45	94	2.1	614	13	BM592900

ALIGNMENTS

RESULT 1
LOCUS AY103544
DEFINITION Zee mays PC0076329 mRNA sequence.
ACCESSION AY103544
VERSION AY103544.1 GI:21206622
KEYWORDS
SOURCE
ORGANISM Zee mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC
Clade; Panicoideae; Andropogoneae; Zee.
REFERENCE
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.

[illegible]

```

Oy 603 ValValTTPserLyaspValLeuThrLeuSerGluGlyAspTyrThrLeuTrrPyrArg 622
    |||
Db 230 ACCGTT-----GATGAGATTAAATCTTACACAGATAGCGTTACGAAATG----- 186
Oy 623 PhseRilLeasnaSnclInlySpGluTrrPlys----- 633
    |||
Db 185 GTTTCAGTACGATTATCCAGAGATGAACTAAATTCGCTGATGACCTCAACATTATGTT 126
    |||
Oy 634 -----LysIleGlySerVal-----SerValYstHrPrThGluTrrThr 647
    |||
Db 125 GTGAGATTAAACAGTGTACTACTACTGAAATGAAATTAATTCGTTAATGAGTCATT 66
    |||
Oy 648 HIsPoleuPhcGluValGlyHIsaSnclIntrSerTrpTyr 661
    |||
Db 65 CATTAATTATGATTAAATGAGATGAGCAGCTGACCTGACCTAT 24
    |||

RESULT 3
AK018613
ACUS
ACCESSION AK018613 2178 bp mRNA linear HTC 19-JAN-2002
VERSION Mus musculus adult male cecum cDNA, RIKEN full-length enriched
KEYWORDS full insert sequence.
SOURCE AK018613.1 GI:12858408
HTC: CAP trapper.
MUS musculus (strain:C57BL/6J) adult male cecum cDNA to mRNA,
clone:J130012D09.
ORGANISM Mus musculus
Mammalia: Eutheria: Chordata: Craniata: Vertebrata: Euteleostomi:
Eukaryota: Euthera: Rodentia: Scurionath: Muridae: Murinae: Mus.
REFERENCE
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE
AUTHORS Carninci, P., Shibata, Y., Hayata, N., Sugahara, Y., Shibata, K.,
Itoh, N., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
JOURNAL Prepare full-length cDNA libraries for rapid discovery of new genes
MEDLINE 20499337
PUBMED 11042159
REFERENCE
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsu, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsuki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
JOURNAL sequencing pipeline with 384 multiplexed sequencer
MEDLINE 20530913
PUBMED 11076861
REFERENCE
AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Aizawa, K., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I.,
Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, M., Gaasterland, T., Gissi, C., Kling, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,
Quackenbush, J., Schriml, L., Staudt, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Bardelli, R., Barish, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M., Brownstein, M., Bult, C.,
Fletcher, C., Fujita, M., Cariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,
Rang, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-Oka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S.,
and Hayashizaki, Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE
AUTHORS 5 (pages 1 to 2178)
Aizawa, K., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
Arakawa, T., Bardelli, R., Bono, H., Brownstein, M., Bult, C.,
Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hasegawa, T.,
Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hirooka, T., Hori, F.,
Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,
Kato, H., Kawaji, K., Kojima, Y., Konno, H., Koyama, S.,
Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,
Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
Tanaka, T., Teijima, Y., Toyo, T., Yamamura, T., Yamana, I.,
Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration and Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in RIKEN
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in RIKEN contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGACGAGATGCGATGATTAATTAATCCCGCCCGCC 3']. cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to Rot -10.0 and subtraction to
Rot -185.2. Second strand cDNA was prepared with the primer
GAGACGAGATGCGATGATTAATTAATCCCGCCCGCC 3'. cDNA was cleaved
with BamHI and XhoI. Vector: a modified plasmid vector KS(+) after
end: BamHI. Host: DH10B.
FEATURES
Source
Location/Qualifiers
1..2178
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="RANTOM:DB:9130012D09"
/db_xref="MGI:1904161"
/db_xref="taxon:10090"
/clone="9130012D09"
/sex="male"
/tissue-type="cecum"
/clone.lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
74..1807
/note="Immunoglobulin domain containing protein
data source: Pfam, source key: PF00047, evidence: ISS
putative"
/codon_start=1
/protein_id="BA31307.1"
/db_xref="GI:12858409"
/db_xref="MGI:191851"
/translation="MELAGFCCSWVILSLPTTPPAPAAHPIHRAIDLSTNOM
ERPLALATLVNOSTVEQREMAVYECNTNNDNITIHVNSLLIVERKLSADN
KTLIILVREDSGSYICEQGEFVORNSNTSLTNYKPPDVSIFDGAAGVAYE
VMGNTVNFVEAQSPPVAYNTLPSDTIOPITGTFTIDAVSRHEGMYRCLVSNP

```

VTNLSRLGVKVOLEKVTAPNIEPTLALVENATSVLTCTKSHORVHMFELKGP
LRSPRLTLEONRTLTHGLORDIDIGPECEVNMWGSQASVPKLTINPDQVEI
TQPRASVSTTEMLNSSLTLCRADSPKARVOTHEHSKVLIDGROLSTEALROE
HOGITCSSTNDVGLASASVLMVLOLSSWSPGALAGIVLILVALALALIGY
FLYSTKDMTRRSASDTSNLTPTVMOSTPESRNKPTVYDNPKEGEGRKG
KMSLIPRSPDQFEKKPPSAAPBGRPRPLQIPKQPLMPPGGNESNVEKLNSN
HSLVCKIRPSA*

BASE COUNT 541 a 631 c 529 g 477 t
ORIGIN

Alignment Scores:

Pred. No.: 0.642 Length: 2178
Score: 111.00 Matches: 79
Percent Similarity: 37.54% Conservative: 55
Best Local Similarity: 22.13% Mismatches: 135
Query Match: 2.308 Indels: 88
DB: 11 Gaps: 21

-030-330-1 (1-843) x AK018613 (1-2178)

488 ThrilethleuLeuTyrArgArgThrGluGlnTTPGluProValArgHisala 507
||||| : : : : :
230 ACCCTCACCA---GTCATCAGACGACCTGTCCACAGAGAGGAATGGCAGTTTCTAC 286
508 GlnGlyGlyTyrValAsnSerIleYsValAsnThrAspProAsnAsnValVal 527
287 TGTACACTANTGCGATACATACATCCATCGGTGCTCCAAACATCTGCTTG 346
528 -----ThyValAspAsnAsnGlnGlyLysLeuSerIleValPro 540
347 TTGATGAGCCATGACCTGTCTGATACATACAGACCTCACCCTCATTTGTCAG 406
541 AsnSerPheValAlaAspLeuAsnSerTyr-----GlnHis----- 552
407 CGGAG-----GACTCGGGGCTTACCTGTGTGAGAGTCAGATGGCTTGAGGTT 457
553 -----SerThrIleThrValGlnPheAsnSerAspSerProAspGlu 567
458 CAGAGAACTACACCTCGCTGCGTGAACATGCTGCTGAT----- 502
568 ArgThrProValAlaPheAlaLeuSerThrGlyAlaThrAlaAspAspValIleSer 587
503 -----CCTGTATCATACAGCTGTGACGACGAGAGTACGCTGGGAGTGGTGAGG 556
588 -----GlyTrpValMetAlaGluValProGlyGlySerSerAsnTyrProVal 603
557 ATGCAAGCAACACCTGACACTTCAAGGTGGAAGACAGATGCTCTGCTGCTGCTAT 616
604 ValTrp-----SerLysAspValLeuThrLeuSerGluGlyAspTyrThrLeu 619
617 GCGTGTATCTCCCGACGACCTTCATCCAGCCCTTACACAGAAATCATCATCAT 676
620 -----TrpTyrArgPheSerIleAsnAsnGlnLysAspGlu 631
677 GCTGTGTCCAGGACATGAGGCGCATGTACAGGCTGCTGCTCCACCTGTCCACAAC 736
632 TrpLysIleGlySerValSerValLysThrProGlnGlyTyrThrHisProLeuPhe 651
737 TTGTCCCGCTAGGTGTGTCAAGTCCAAAGCTTGAAGAGTACGTCCAGGAATATC 796
652 Glu-----ValGlyHisAsnGlnThrSer-----ThyTyrThrLeuAspMet 665
797 GAGTCCCAACCTTGCGCTTGTGGAATGCGACCTGTGTGACACTGACCTGCAAAAC 856
666 AlaHisAsnArgValLeuProAspPheThrLeuLys----- 677
857 AGCCACCAAGAGGCTGTGCTCATGTTCTAAAGGTCACACCACTCAGGCTAGTGAC 916
678 AsnLeuGlyLeuProPheAsnGlyLysLeuValValPheArgGlnThrInserSer 697
917 GCGCTGACACTGTATCCCAAAACAGAACCTGACATCCAGCTTCAGAGATGAT 976
698 SerGlySerLeu-----TrpAlaAlaGln-----GluThrValHisIle 710

DB 977 ATTGGCCCTATGAGCTGAGCTGAGACTGGGGATAGCCAGCAGCGCTCCCTG 1036
OY 711 LysGlnGlyGluThrPheValTyrIlePro-----ValValGluGlyProIle 726
DB 1037 AAGCTC-----ACCATCAACTATGGCCCCGACCAAGTGAATCATCCAGGGGCCAGG 1090
OY 727 ProAspGlySerTyrArgAlaThrLeuHisAlaPheValAsnGlyGlnGlnGlnLeuTyr 746
DB 1091 -----TCTGAGGTGTGACACTATTTGAGCAATGCTCACTCCAGCTTGACCTGTAT 1144
OY 747 -----LeuLysGlyLysArg-----AsnTyrThrValLys 756
DB 1145 TCGCCGGCGAGCTCCATACACAGCTGCGCAGCTATCACTGACCATATGACACTCTTAAG 1204
OY 757 ILeValAsnGly-----ThrAlaValGlnAlaIleGluSerSerGlnGlu 771
DB 1205 GTGCTTGATGGGAGCAGCTGAGCATCGAAGCACTGAGCAGAGCATCATCAG 1255

RESULT 4

BM468550/c 1093 bp mRNA linear EST 05-FEB-2002
LOCUS
DEFINITION AGENCOURT_6432286 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5535892
5', mRNA sequence.
ACCESSION BM468550
VERSION BM468550.1 GI:18517592
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1093)
AUTHORS NIH-MGC http://mgi.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgrabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM12224 row: 1 column: 05
High quality sequence stop: 644.
Location/Qualifiers
1. 1093
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5535892"
/clone_1lb="NIH_MGC_71"
/tissue_type="leiomysarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-Sport6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo df.
Average insert size 2.1 kb."

FEATURES

source

BASE COUNT 254 a 304 c 257 g 277 t 1 others
ORIGIN

Alignment Scores:

Pred. No.: 1.15 Length: 1093
Score: 104.00 Matches: 79
Percent Similarity: 33.14% Conservative: 36
Best Local Similarity: 22.77% Mismatches: 120
Query Match: 2.34% Indels: 113
DB: 13 Gaps: 19

US-10-030-330-1 (1-843) x BM468550 (1-1093)

OY 116 MetProAspAsnLeuArgMetTrpLeuGlnIleTyrAspGlnGluIleGlyLeuLeu 135
DB 903 TTGCCCAAGTAGATTCGATTTTGGCTCAAG----- 874

QY	136	SerGlyLyAlaGlnLeuSnsGlnGluLeuLeuArgThrGlnGlyValProAlaGluVal	155
Db	873	-----GCAGATTTTATACCACCGGGTGT-----	847
QY	156	HisAlaLeuMetAspAsnGlyHisPheAlaSnsSprMetLarGTrpAsnGlnGlyTr	175
Db	846	---GCCAGCTTTTCGAAAGGAGGTTT-----TAAGGTAT	814
QY	176	-----ProTrpAsnAsnLysGluProLeuLeuProAsnGlyAsnHisAlaTrpThr	192
Db	813	CTTTCCCTCCCTGGTGGTGGCTTGGCCGATTGGTCGATAGAGGCTCCAGACNACAC	754
QY	193	GlyCysValAlaTrp-----AlaAlaAlaGlnLeuMetLarGTrpHisSer	207
Db	753	AGTTAGTAGCAGACCCCAAGATTTAGAAAGCTCCCTCCATCCAAAGATACGAAGA	694
QY	208	TrpProLeuGlnGlyLeuGlySerPheAspTrpHisAla-----GlySerLe	223
Db	693	AAACCTGATTTAGAGGGGGCCAGAGTACAGAGAACAGAGGGCTGCTGCATAGGCCATC	634
QY	223	uValGlyAsnTrpSer-----GlyThrPheGlnGlyMetLarTrpAspTr	237
Db	633	CGTTGGCCAAATGGCGCTCTGTAGAAATACTTGCCAATCATGGTTCAG-----AGCTG	580
QY	237	PLeuSnsMetProGlyAsnProAsnLeuThrGlnSerGlnValAlaAla--	256
Db	579	GATTCAAGTTGAAAGCCAGAACCCAGACGTCCTCAAT-----GACTCCATTAACAGCGATA	526
QY	257	-----TyrAlaThrLeuMetLarGlnPheValSerAlaSerValSerMetSerPheTrg	274
Db	525	GTTGGCTGTACGGTGTATACAGAAAGAGCGCAATCAACAGCCACACCGCTGACA-----	474
QY	274	uAsnGlySerGlyThrTrpSerValTrpValValGlyAlaLeuArgSnsAsnPheArgTr	294
Db	473	-----GCACCCCAACCAATCATGACGTA	451
QY	294	LysArgSerLeuGlnLeuHisValArgAlaLeuTrpThrSerGlnGlnTrpHisAspMe	314
Db	450	TACGACTTCATATTTATTTATTCACAGAGATCTGTATGAGCACCTCCACCATATGATAC	393
QY	314	fileArgGlyLeuAlaSerGlyArgProValTrpTrpAlaGlyAsnAsnGlnSerIl	334
Db	392	-----AGCCAAACATTTATGAGAGATATCCATGACACTCGTGGTGGCATGGAATA	340
QY	334	egLy---HisAlaPheValLysAspGly-----TyrAlaSerAs	346
Db	339	TACCTTCATCTCTTGAAGAAGACAGTAAATGATTTGGAAATGATAAAGAAAGCCGA	280
QY	346	pGlyThrPheHisPheAsnTrpGly-----TrpGlyGlyValSe	359
Db	279	TGGAAATGAAGCTTCATTAACCTCCAGCCCAAGCTCCATCTGAGGGGGGTGATTC	220
QY	359	rAsnGlyPheTrpLysLeuThrLeuLeuSerProThrSerLeuGlyLe-----	375
Db	219	CAACTGGA-----CCTCCAACTGCTGGTGGCATTTTAAGTGTAG	181
QY	376	-----GlyGlyGlyGlyLeuGlyPheThrLysLeuGlnGluLe-----	388
Db	180	ATAAATTCAAAAATGAGCTCAATTCCTCAATATCTGTTATATCGGAGATCCGAGGCC	124
QY	389	-----IleThrGlyIleGluProAlaLysThrProAlaGluAlaLysThrAspA	405
Db	123	CGAGCGTGGCGAGCGTGGTGGCGAGCTGAGATTCCAGGGGGCCAGAGTCCGGCT	65
QY	405	laleuProIleLeuAla	410
Db	64	CCCGCTCTTCGCTCGCT	48
QY	4004957	AK004957	3170 bp mRNA linear HTC 19-JAN-2002
DEFINITION			Mus musculus adult male liver cDNA, RIKEN full-length enriched

QY	136	SerGlyLyAlaGlnLeuSnsGlnGluLeuLeuArgThrGlnGlyValProAlaGluVal	155
Db	873	-----GCAGATTTTATACCACCGGGTGT-----	847
QY	156	HisAlaLeuMetAspAsnGlyHisPheAlaSnsSprMetLarGTrpAsnGlnGlyTr	175
Db	846	---GCCAGCTTTTCGAAAGGAGGTTT-----TAAGGTAT	814
QY	176	-----ProTrpAsnAsnLysGluProLeuLeuProAsnGlyAsnHisAlaTrpThr	192
Db	813	CTTTCCCTCCCTGGTGGTGGCTTGGCCGATTGGTCGATAGAGGCTCCAGACNACAC	754
QY	193	GlyCysValAlaTrp-----AlaAlaAlaGlnLeuMetLarGTrpHisSer	207
Db	753	AGTTAGTAGCAGACCCCAAGATTTAGAAAGCTCCCTCCATCCAAAGATACGAAGA	694
QY	208	TrpProLeuGlnGlyLeuGlySerPheAspTrpHisAla-----GlySerLe	223
Db	693	AAACCTGATTTAGAGGGGGCCAGAGTACAGAGAACAGAGGGCTGCTGCATAGGCCATC	634
QY	223	uValGlyAsnTrpSer-----GlyThrPheGlnGlyMetLarTrpAspTr	237
Db	633	CGTTGGCCAAATGGCGCTCTGTAGAAATACTTGCCAATCATGGTTCAG-----AGCTG	580
QY	237	PLeuSnsMetProGlyAsnProAsnLeuThrGlnSerGlnValAlaAla--	256
Db	579	GATTCAAGTTGAAAGCCAGAACCCAGACGTCCTCAAT-----GACTCCATTAACAGCGATA	526
QY	257	-----TyrAlaThrLeuMetLarGlnPheValSerAlaSerValSerMetSerPheTrg	274
Db	525	GTTGGCTGTACGGTGTATACAGAAAGAGCGCAATCAACAGCCACACCGCTGACA-----	474
QY	274	uAsnGlySerGlyThrTrpSerValTrpValValGlyAlaLeuArgSnsAsnPheArgTr	294
Db	473	-----GCACCCCAACCAATCATGACGTA	451
QY	294	LysArgSerLeuGlnLeuHisValArgAlaLeuTrpThrSerGlnGlnTrpHisAspMe	314
Db	450	TACGACTTCATATTTATTTATTCACAGAGATCTGTATGAGCACCTCCACCATATGATAC	393
QY	314	fileArgGlyLeuAlaSerGlyArgProValTrpTrpAlaGlyAsnAsnGlnSerIl	334
Db	392	-----AGCCAAACATTTATGAGAGATATCCATGACACTCGTGGTGGCATGGAATA	340
QY	334	egLy---HisAlaPheValLysAspGly-----TyrAlaSerAs	346
Db	339	TACCTTCATCTCTTGAAGAAGACAGTAAATGATTTGGAAATGATAAAGAAAGCCGA	280
QY	346	pGlyThrPheHisPheAsnTrpGly-----TrpGlyGlyValSe	359
Db	279	TGGAAATGAAGCTTCATTAACCTCCAGCCCAAGCTCCATCTGAGGGGGGTGATTC	220
QY	359	rAsnGlyPheTrpLysLeuThrLeuLeuSerProThrSerLeuGlyLe-----	375
Db	219	CAACTGGA-----CCTCCAACTGCTGGTGGCATTTTAAGTGTAG	181
QY	376	-----GlyGlyGlyGlyLeuGlyPheThrLysLeuGlnGluLe-----	388
Db	180	ATAAATTCAAAAATGAGCTCAATTCCTCAATATCTGTTATATCGGAGATCCGAGGCC	124
QY	389	-----IleThrGlyIleGluProAlaLysThrProAlaGluAlaLysThrAspA	405
Db	123	CGAGCGTGGCGAGCGTGGTGGCGAGCTGAGATTCCAGGGGGCCAGAGTCCGGCT	65


```

OY 443 GYTYFARGLeuAnLysAlaAspGlyValIleGluValLysThrSerSerIleAsn 462
Db 1746 -----GAC 1748
OY 463 ILESerTPTYGlyTYGlyGluHisProGluSerPheSerLeuAlaProAsnGlnLeu 482
Db 1749 TTCTCGTGTGGATGATCCCTTCAC----- 1775
OY 483 SerGlnGlyIleAsnThrIleThrLeuLeuTYFARGThrGlyThrGlnGlnTrpGlu 502
Db 1776 -----GGGTCACAGTGGCAACACAGACCTCTCTGATTTGGAGCCGACCTGGAA 1829
OY 503 ProVal---ArgHisAlaGlnGlyTYFValAsnSerIleLysValAsnThrThrAsp 521
Db 1830 AATGTCAGCAGATGACCGCTCACACCAACAAAG----- 1865
OY 522 ProAsnAsnValValValThrValAspAsnAsnGlnGlyLysLeuSerIleValProAsn 541
Db 1866 -----AACCAAGAGGAAAGAGCTGGGAAAGGTGTACGGC 1901
OY 542 SerPheValAlaAspLeuAsnSerTYGlyGluHisSerThrIleThrValGlnPheAsnSer 561
Db 1902 TACTTCTCCCGAAC-----CGCCAGACGACATCTACTATC----- 1937
OY 562 AspSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrGly----- 578
Db 1938 ---TCGGAGATTAAGCGCATGGGAAAGCTGGGTACTCTCTGCAAGTGTATGTGAGA 1994
OY 579 -----AlaThrAlaAspAspValIle 585
Db 1995 GTCAATGGACCTTAACCAAGCTCGCTGGTGGAGCCCTCATGATGACGTGTCT 2054
OY 586 SerLeuGlyTYFValMetAlaGluValProGlyLysSerSerAsnTYFProValValTrp 605
Db 2055 AAAATGGCATTCCTA----- 2069
OY 606 SerLysAspValLeuThrLeuSerGlnGlyAspTYFThrLeuTYFArgPheSerIle 625
Db 2070 -----ACCATGACCTGCGACAGGCTGGAGCCGATCGATGACCACTAGCT--- 2117
OY 626 AsnAsnGlnLysAspGluTrpLysLysIleGlySerValSerValLysThrProThrGlu 645
Db 2118 -----CCCGAG 2123
OY 646 TYFThrHisProLeuPheGluValGlnHisAsnGlnThrSerThrTYFThrLeuAsnMet 665
Db 2124 AACACACAGCCTCTCTG-----CTGACGACATTCAGTGGAGACCGA 2165
OY 666 AlaHisAsnArg-----ValLeuProAspPheThrLeuLysAsnLeuGlyLeuPro 682
Db 2166 CGCTTTCTCGATTTGGTAGTGTTCGACCTGACTGACCTGGATGATGATGCTTA--- 2222
OY 683 PheAsnGlyLysLeuValValVal-----PheArgGlnThrGlnLysSerSerGlySer 700
Db 2223 -----GATGAATTCATCATGGCTGCCCACTGACAGATACAGATGTCACCTTCCGAGCTG 2276
OY 701 LeuTrpAlaAlaGlnGlnLysValHisIleLysGlnGly 713
Db 2277 CTAGGGGAGAGAGATGGCGAGTGTACATTATTAATAGGC 2315

```

```

RESULT 6
LOCUS AU140457 820 bp mRNA linear EST 05-AUG-2002
DEFINITION AU140457 PLACE3 Homo sapiens cDNA clone PLACE3000100 5', mRNA
sequence.
ACCESSION AU140457.1 GI:11001978
VERSION AU140457.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 820)

```

```

AUTHORS Ota,T., Suzuki,Y., Saito,K., Ishii,S., Yamamoto,J., Sugiyama,T.,
TITLE Nishikawa,T., Nakamura,Y., Sugano,S., Masuhara,Y. and Isogai,T.
HRI human cDNA project (Ota,T., Suzuki,Y., Saito,K., Ishii,S.,
Yamamoto,J., Sugiyama,T., Nishikawa,T., Nakamura,Y., Sugano,S.,
Masuhara,Y., Isogai,T.)
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source
1. 820
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE3000100"
/clone_1lb="PLACE3"
/issue_type="Placenta"
/issue="Vector: PME18PL3"
/location="PME18PL3"
/location/Qualifiers
BASE COUNT 218 a 220 c 183 g 195 t 4 others
ORIGIN
Alignment Scores:
Pred. No.: 1 Length: 820
Score: 102.50 Matches: 58
Percent Similarity: 37.41% Conservative: 46
Best Local Similarity: 20.86% Mismatches: 89
Query Match: 2.31% Indels: 85
DB: 9 Gaps: 13
US-10-030-330-1 (1-843) x AU140457 (1-820)
OY 412 LysAspArgIleGluAlaGluTrpLysSerGlnSerGlyLeuAsnValGly----- 427
Db 82 AGAGGTGTCATTTTCCTCCAGACACATGACAGTGGCTTACAAATTCCTCTGCTGATA 141
OY 428 -----TySerIleTyAsn-----ThrGly 434
Db 142 GATATGTAGTGAGTCTCTCCAGTGTCTGCAACAAACATGAGACACACCTTGTAGAGA 201
OY 435 GlnGlnGlnSerAsnLysAsnLysGlyTYFArgLysAsnLysAlaAspGlyValIle 454
Db 202 AGACGAAACACAGCTTATATCCCACTGGCATGACTTTTCTGAT----- 249
OY 455 GluValLysThrSerSerIleAsnIleSerTrp----- 465
Db 250 ---ATTACGCCAACCTTTTACTGTGCACTGGATTTGCTTCGAGCCACATCACTGCG 306
OY 466 TYGlyTYGlyGlnHisProGluSerPheSerLeuAlaPro-----AsnGlnLeuSer 483
Db 307 TACAGGATGGCCATCATCCGACGACCTTACATGGGAGACCTCGAGAAATGGGGTCCC 366
OY 484 GlnGlyIleAsnThrIleThrLeuLeuTYFArgArgTYFThrGlnGlnGlnTrpGlu 503
Db 367 CACTCTCGGAATTCACACCTGACCAACCTCACTCCAGCAGAG----- 414
OY 504 ValArgHisAlaGlnGlnGlyTYFValAsnSerIle---LysValAsnThrThrAspPro 522
Db 415 -----TATGTGTGACAGCATGTGTGCTTATATGACAGAGAGAA 453
OY 523 AsnAsnValValVal-----ThrValAspAsnAsnGlnGlyLysLeuSerIle 538
Db 454 AGTCCCTATATGATTTGGCAACATCAACAGATTTTGATGTTCCGAGGACCTGGAAATT 513
OY 539 ValProAsnSerPheValAlaAspLeuAsnSerTYGlyGlnHisSerThrIleThrValGln 558
Db 514 GTTGTGTCAGCCGCCACGACCTTACTGATCACTGGAGTCTCTGCTGTACAGTGA 573

```

QY 559 PheAsnSerAspSerProAspGluLeuArgThrProValAlaPheAlaLeuSerGlnGly 578
Db 574 TAT-----TACAGGACTTTACGGA 594
QY 579 AlaThrAlaAspAspValIleSerLeuGlyTrpValMetAlaGluValProGlyGlySer 598
Db 595 GAACA-----GGAGGAAT 609
QY 599 Ser-----AsnTyrProValValTrpSerLysAspValLeuThrLeuSer----- 613
Db 610 AGCCCTGTCAGAGTTCACCTGCTGGAGCAGCACTACACCTACACCTACAGGCTT 669
QY 614 -----GluGlyAspTyrThrLeuTrpTyrArgPheSerIleAsnGlnLysAspGlu 631
Db 670 AAACCTGGAGTGTATATACATC-----ACTGNGTATGCTGTCTACCTGGCCCTGGAGCAGC 726
QY 632 TrpLysLysIleGlySerValSerValLysThrProThrGluTyrThrHisPro 649
Db 727 CCCGACGACGACGACCAATTTCCATTATACCGAACAAGAAATTTGGCAACCA 780
QY 727 BH770545/c 1594 bp DNA linear GSS 01-MAY-2002
LOCUS LMGtag308 MG1363 Random Sequence Tag Library Lactococcus lactis
DEFINITION subsp. cremoris genomic, DNA sequence.
ACCESSION BH770545
VERSION BH770545.1 GI:20373502
KEYWORDS GSS.
SOURCE Lactococcus lactis subsp. cremoris.
ORGANISM Lactococcus lactis subsp. cremoris
Bacteria: Firmicutes; Lactobacillales; Streptococcaceae;
Lactococcus.
1 (bases 1 to 1594)
Boltin, A., Ehrlich, S.D. and Sorokin, A.
Studies of genomes of dairy bacteria Lactococcus lactis
Sci. Aliments, (2002) In press
Contact: Sorokin A
Genetique Microbiome
INRA
CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France
Tel: 33 1 34 65 25 16
Fax: 33 1 34 65 25 21
Email: sorokine@jouy.inra.fr
best homologue in strain IL403 is ynfH (27%)
Class: shotgun
High quality sequence start: 30
High quality sequence stop: 1566.
Location/Qualifiers
1. 1594
/organism="Lactococcus lactis subsp. cremoris"
/strain="MG1363"
/db_xref="taxon:1359"
/clone_lib="MG1363 Random Sequence Tag Library"
/note="Vector: pSGM02; Site.1: SmaI; Library of
chromosomal fragments of L.lactis strain MG1363 was
prepared by partial AluI digestion or by sonication."
BASE COUNT 362 a 354 c 397 g 466 t 15 others
ORIGIN
Alignment Scores:
Pred. No.: 3.53 Length: 1594
Score: 102.50 Matches: 124
Percent Similarity: 31.70% Conservative: 64
Best Local Similarity: 20.91% Mismatches: 196
Query Match: 2.31% Indels: 209
DB: 17 Gaps: 28
US-10-030-330-1 (1-843) x BH770545 (1-1594)
QY 271 SerPheTyrGluAsnGlySerGlyThrTyrSerValTyrValGlyAlaLeuArgAsn 290
Db 1535 TCTCTTGGCGCGATGNAAGAAAGAGCNCNATATATGCGTGTCTTTGGCCCTCCCAATN 1476

QY 291 AsnPheArgTyrLysArgSerLeuGlnLeuHisValArgAlaLeuTyrThrSerGlnGlu 310
Db 1475 NATTTCTCCSAGAGCTCCNCATGTGA-----AGTCTCTAAGTAATAGCCCAAT 1425
QY 311 TrpHisAspMet-----IleArg----- 316
Db 1424 TGGTATTCGGTTCAGGAGACTCCCGTCCATTTNCCCNNTCCANATTAAGATTAATAC 1365
QY 317 GlyGluLeuAlaSerGlyArgProValTyrTyrAlaGlyAsnAsnGln----- 332
Db 1364 GGAGATCCAGTTGTGAGCAGGAGCCACAGATGAAGTGAATTAACAGTACGTCTCCCA 1305
QY 333 -----SerIleGlyHisAlaPheValLysAspGlyTyrTyrLysSerProGlyThrPheHisPhe 351
Db 1304 GGCTCAGTTGTCTTAATGCTCCATTCCTCAATTCACACCAACAGATTAAGTTAGTTCA----- 1254
QY 352 AsnTrpGlyTrpGlyGlyValSerAsnGlyPheTyrLysLeuThrLeuLeuSerProThr 371
Db 1253 -----GATATGTCAGTGAACCCACCTACAGCAGCAACCCAGACATCCAACT 1206
QY 372 SerLeuGlyTyrLeuGlyGlyGlyIleGlyPheThrIleTyrGlnGluIleIleThrGly 391
Db 1205 CTAGTAGCT-----CCAAACAGT 1188
QY 392 IleGluProAlaLysThrProAlaGluAlaGlyThrAspAlaLeuProIleLeu----- 409
Db 1187 AACTTACCTGCGAACCTTCAGCTTGAGCCGACGACATTCCTGCTTCCAGCG 1128
QY 410 AlaLeuLysAspIleGluAlaGlyTyrLysSerGlySerGlyLeuAsnValGlyTyrSer 429
Db 1127 AGCTCTCAAGAC----- 1116
QY 430 IleTyrAsnThrGlyGlnGlnSerAsnLeuAspLeuGlyTyrArgLeuAsnLysAla 449
Db 1115 -----AGTACAGCA-----GCT 1104
QY 450 AspGlyGluValIleGluValLysThrSerSerIleAsnIleSerTrpGlyTyrGly 469
Db 1103 GACATTCACCTCACTCAGTGAATTCACATTC----- 1071
QY 470 GlnHisProGluSerPheSerLeuAlaProAsnGlnLeuSerGlnGlyIleAsnThrIle 489
Db 1070 -----TCAGGACAGCAGTAACCAATGTAACCTATACAGGGTACTACACGCTC 1020
QY 490 ThrLeuLeuTyrArgArgThrGlyThrGlnGlnTrpGluProValArgHisAlaGlnGly 509
Db 1019 ACT-----TATAGCGCGCAGCA-----TATGACAGCTGACCCAAACG----- 981
QY 510 GlyTyrValAsnSerIleLysValAsnThrThrAspProAsnAsnValValAlaThrVal 529
Db 980 -----CTGATTTGCTCTGACCCAGCAGATACACAGTCCCAAT-----GCTCCAAACGTT 930
QY 530 AspAsnAsnGlnGlyLysLeuSerIleValProAsnSerPheValAlaAspLeuAsnSer 549
Db 929 GGAATGTCACAGGAATTAACAC-----AATGCTATACGGTTACAGGAGCAGCC 879
QY 550 TyrGlnHisSerThrIleThrValGlnPheAsnSerAspSerProAspGluIleArgThr 569
Db 878 GAACCTAATTCACACATCAATTAAGATTAATATGAGAT----- 837
QY 570 ProValAlaPheAlaLeuSerThrGlyAlaThrAlaAspAspValIleSerLeuGlyTyr 589
Db 836 -----ACGATGGAAACCGGACGACAGCAGCA----- 810
QY 590 ValMetAlaGluValProGlyGlySerSerAsnTyrProValValTrpSerLysAspVal 609
Db 809 -----AGTGAATTAACACAGTACTCTCCAGGCTCAGTT 774
QY 610 -----LeuThrLeuSerGlyLysAspTyrThrLeuTyrArgPheSer 624
Db 773 GGTCTCAGCTCCATTCACATCACTGACGACAGAT----- 738

```

OY 625 ILeasnaSngLlyAspGluTrpLysLysIleGlySerValSerValLysThrProThr 644
DB 737 -----GTTCAAGTCATGTCAGTACCAAC 711
OY 645 GluTrpThrHisProLeuPheGluValIly-----HisAsnGln 657
DB 710 TCAGGACCAATCTCCAGACACCACTAGTCCGCTTTAGTGGCCCAACAGATATTTA 651
OY 658 ThrSerTrpTrpThrLeuAspMetAlaHisAsnArgVal-----LeuProAspPheThr 675
DB 650 ACTGGGACCACTCTTAAAAAGGTCGACTTGCACCAACAGTTAGTCTACTCT--GCTACA 594
OY 676 LeuLysAsnLeuGlyLeuProPheAsnGlyGluLeuValValIlePheArgGlnThrGln 695
DB 593 CTTAAAGAT-----AGTGAAGTTAAACTGTCGCGGTACTCATGTATTATACC 546
OY 696 SerSerSerGlySerLeuTrpAlaLalaGlnGluThrValHisIleLysGlnGlyLthr 715
DB 545 AACTCAGCAGCAATGTA-----528
OY 716 PheValLysTrpProValValGlyGlyProIleProAspGlySerTrpArgAlaThrLeu 735
DB 527 -----GTACCGAATGCGCAACTGTCGCGGTACTCATGTATTAT 483
OY 736 HisAla-----PheValAsnGlyGlnGlnLeuTrpLeuLysGlyLysArgAsnTrp 753
DB 482 TCAGCAGAGGATATGCAATATGACTCAACCTGTAGTGTGATGTCAGTGTCCACCAACCA 423
OY 754 ThrValLysIleVal-----AsnGlyThrAlaValGluAlaIleGlySer 768
DB 422 GCTAGCCCTCTCTTAGTAGCCCAACAGATATTATTAACCTGCACTCTTAAAAAGGT 363
OY 769 SerGluGluLysArgValPheProAsnProAlaArgAspTrpValIleLysSerAlaPro 788
DB 362 GCAGCTGAC-----GCAACAGCTAGTACTACT 336
OY 789 CysIleProGlnLthr-----SerIleLeuLeuPheAsnLeu 801
DB 335 GCTACACTTAAACATGATGTATGTTAACTGTTTCGCTACTCATGTATTATCAACACTCA 276
OY 802 SerGlyLysIleValIleMetLysAsnSerLeuSerAlaGly 814
DB 275 GCAGGAATGTTAGTACCAAGTGCACAACTGTCGCGAGT 237

RESULT 8
LOCUS BG245705 912 bp mRNA linear EST 13-FEB-2001
DEFINITION 602358489p1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4486942 5',
G3'
FEATURES
SOURCE: house mouse.
ORGANISM: Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 912)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaaps-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
http://image.llnl.gov
Plate: LHM10331 row: a column: 23
High quality sequence stop: 703.
Location/Qualifiers
1..912

```

```

OY 482 LeuSerGlnGlyIleAsnThrIleThrLeuLeuTrpArgArgThrGlyThrGluGlnTrp 501
DB 56 CTCAGCACTGGCTCCACAGTAACTATCTCTC-----AGGGTACGACCAACCTCG 109
OY 502 GluProValArgHisAlaGlnGlyLysTrpValAsnSerIleLysValAsnThrAsp 521
DB 110 ATTCAGAGAGAGCCCATGCTGGC-----GTTATGGCTGGATGTGACAGCTGAGC 163
OY 522 ProAsnAsnValValValThrValAspAsnGlnGlyLysLeuSer-----Ile 538
DB 164 CCTTTGATCTTGTTCATACCCCTTCATCATCAAGAAAGGGGAGATCACTGGGAAGTTCCA 223
OY 539 ValProAsnSerPheValAla-----AspLeuAsnSerTrpGlnHisSerThrIleThr 556
DB 224 ATGCCCTCAGCAGAGGTGGCCAGCCCTTCATCATCTGATTAACAAGATGCGACTGTACT 283
OY 557 ValGlnPheAsnSerAspSerProAspGlu-----IleArg 568
DB 284 GTACGTTAC-----TCACCAAGTAAGAGCGCCCTGATGAATGCAATTCGCTGTAT 334
OY 569 -----ThrProValAlaPheAlaLeuSerThrGlyAlaThrAla 581
DB 335 GACATATGCAATATCCAGGAAGCCCTCTGCACTTATCTATCT-----376
OY 582 AspAspValIleSerLeuGlyTrpValIleValGluValProGly-----GlySer 598
DB 377 ---GATTATGTCAACAGTGTGCCACATCACTGCTTATGCTGTGGCTTACCATGAGAGTG 433
OY 599 SerAsnTrpProValValTrpSerLysAspValLeuThrLeuSerGluLysAspTrpThr 618
DB 434 GTCACAAACCTGCCACCTCTGATCAACACCAAGATGACAGAGAGGGGCGTTG---490
OY 619 LeuTrpTrpArgPheSerIleAsnGlnLysAspGluTrpLysLysIleLysVal 638
DB 491 -----TCTCTG 496
OY 639 SerValLysThrProThrGluTrpThrHisProLeuPheGluValGlyHisAsnGlnThr 658
DB 497 GCCATTGAGGCGCATCTTAA-----GCAGAAATGAGTTGCACATGACAC 541
OY 659 SerTrpTrpThrLeuAspMetAlaHisAsnArgValLeuPro---AspPheThrLeu---676
DB 542 CAGGATGCAACATGCAAGTGTCTTACCTGCTGTACTGCTGTGACTATAGCATCTTA 601
OY 677 -----LysAsnLeu---GlyLeuProPheAsnGlyLys---LeuValVal 689
DB 602 GTTAGTCAATGATCAACATCCAGGACAGTCCCTTACTGCAAGATTAACAGTGAAG 661

```


QY 690 ValPheArgGlnThrGlnSerSerGlySerLeuTrpAlaGlnGlnThrValHis 709
 Db 662 ATATGCGGATATGTC-----CAC 679
 QY 710 IleValGlnGlyGlnThrPheValTrpLysProValValGlnGlyProIle----- 726
 Db 680 CTAAAGGTGGGT-----CCTGCTGCTGATATCCCATCAATCAATCTCA 721
 QY 727 ----ProAspGlySerTrpArgAlaThrLeuHisAla-PheValAsnGlyGlnGlnGln 745
 Db 722 GAACAGACCTTACCTTCCAGACCTGGGTGCCCCCTC----- 761
 QY 745 uTrpLeuLysGlyLysArgAsnTrpThrValLysIleValAsnGlyThrAlaVal 763
 Db 762 -----GGTGAAGAACCTGTCTCTTTAAAGTTGCAGACGCCGGGGAT 806
 RESULT 9
 BG251484 730 bp mRNA linear EST 13-FEB-2001
 602363943F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4472074 5',
 mRNA sequence.
 BG251484 1 GI:12761300
 EST.
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 730)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLM10292 row: f column: 11
 High quality sequence stop: 599.
 Location/Qualifiers
 1.730
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="4472074"
 /clone_lib="NIH-MGC_90"
 /tissue_type="adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: liver; Vector: PCMV-SPORT6; Site: 1; Note: Site_2: SalI; Cloned unidirectionally; Oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
 BASE COUNT 185 a 194 c 199 g 152 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.2 Length: 730
 Score: 101.00 Matches: 53
 Percent Similarity: 37.44% Conservative: 23
 Best Local Similarity: 26.11% Mismatches: 69
 Query Match: 2.28% Indels: 58
 DB: 12 Gaps: 9
 US-10-030-330-1 (1-843) x BG251484 (1-730)
 QY 342 GYTtYAlAsErSpLyThrPheHisPheAsnTrpGlyTTrpGlyValSerAsnGly 361
 Db 226 GGTTTCCTCTCCATGGA-----CACAGTCACAGAGGCTGGATGCGCATACATGCG 279

QY 362 PheTrpLysLeuThrLeuLeuSerProThrSerLeuGlyIleGlyGlyIleGly 381
 Db 280 GCTTACAGGCGGATTTGCTTCCATCCACCACTGACCTGATCCAGGCCAATGA 339
 QY 382 PheThrIleTrpGlnGlnIleIleThrGlyIleGlu-----ProAlaLys 396
 Db 340 AGTTCCGAAGTGAAGAAATATGACAGGATTCATCAGACACCGCTGCTCAGCCC 399
 QY 397 ThrProAlaGluAlaGlyThrAspAlaLeuProIleLeuAlaLeuLysAspIleGluAla 416
 Db 400 CAAGATCGCGTGGGATGT-----TTT 423
 QY 417 GluTrpLysSerGlnSerGlyLeuAsnValGlyTrpSerIleTrpAsnThrGlyGlu 436
 Db 424 GAGGCCAAGGCCCGCATGGTTTC--TGGGATTCCTCAATCAGACAGACGCGGATG 480
 QY 437 GlnSerAsnLeuAspLeuGlyTrpArgLysAlaLysAlaAspGlyGluValIleGluVal 456
 Db 481 GGGAGCGCGCTG--GTGGGCATCATCTC-----CTCCAGGACAAATGATTTCTCA 531
 QY 457 LysThrSerSerIleAsnIleSerTrpTrpGlyTrpGlyLysHisProGlnSerPheSer 476
 Db 532 AAGAGGAGGACATGACTGTTCTTGGACAGATATATGACCAAGGAAAGACTTGGT 591
 QY 477 LeuAlaProAsnGlnLeuSerGlnGlyIleAsnThrIleThrLeuLeuTrpArgThr 496
 Db 592 GGTAGCCCC-----CTGCAGGCGCATCACACT----- 618
 QY 497 GlyThrGlnGlnTrpGluProValArgHisAlaGlnGlyGlyTrpValAsnSerIleLys 516
 Db 619 -----CGACGCGCGCCACATGAATTTCTCAG 645
 QY 517 ValAsnThrThrAspProAsnValValThrValAspAsnAsnGlyGlyLysLeu 536
 Db 646 GCG-----CAGCAGAGGAGGAAGTG 666
 QY 537 SerIleVal 539
 Db 667 GCCATGTGA 675
 RESULT 10
 AK018522 2530 bp mRNA linear HTC 19-JAN-2002
 LOCUS
 DEFINITION
 Mus musculus adult male colon cDNA, RIKEN full-length enriched library, clone: 9030421L11: similar to MEMBRANE GLYCOPROTEIN, full insert sequence.
 ACCESSION
 AK018522
 VERSION
 AK018522.1 GI:12858263
 KEYWORDS
 HTC; CAP trapper.
 SOURCE
 Mus musculus (strain: C57BL/6J) adult male colon cDNA to mRNA, clone: 9030421L11.
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.
 REFERENCE
 AUTHORS
 Carninci, P. and Hayashizaki, Y.
 TITLE
 High-efficiency full-length cDNA cloning
 JOURNAL
 Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE
 99279253
 PUBMED
 10349636
 TITLE
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL
 Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE
 20499374
 PUBMED
 11042159
 REFERENCE
 AUTHORS
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,


```

Db      324 TGCAAGCCATCATGCTCTACAGCAATAGCGTTTCCCAATGAAGAACTTCAGCA 383
Qy      485 -----GlyIleAsnThrIleThrIleuLeuTyraArgArgThGlyThrIuGlnTrpGluPr 503
Db      384 GCTGCATTTAAATACGCGAGCGCTTTGTGT-----GACTGCCAGCTGAGATGGCTCC 437
Qy      503 ovalaRgHlaIagInglYgIyValaAsnSerIleValaAsnThrThrAspProAs 523
Db      438 ACAGTGGGTGGCGGAGAACACTTTCAGACTTCGTAATGCCAGTTGTGCCACCTCA 497
Qy      523 nasnValValaThrValaAsnAsnInglYgIyLysSerIleValaProAsnSerPh 543
Db      498 GCTGCT-----AAAGGAGAGAGCATTTTCAGCTGACGCCAGATGGCTT 542
Qy      543 eValaIaAspLeuAsnSerTyrgIuHISerThrIleThrValaGlnPheAsnSerAsp 563
Db      543 TGtGTGTGAT-----GATTTCGCAAAACCCAGATACCGCTCCAGAGAAAGGAGCTC 596
Qy      563 rProAspGluIleArGThrProValaIaPheAlaLeuSerThrgIyAlaThrAlaAsp 583
Db      597 GCGAATAAAG-----GCTCCGACGCTGAGTTTACTGCTGCTGCTGACAGCACTGACTC 653
Qy      583 pValIleSerLeuGlyTrpValMet-----AlaGluValProG1 596
Db      654 CCTATAGACCTGCTGCTGGAAGAAGAACAGAACGCCCTTCAGAGTACAGATG----- 708
Qy      596 yGlySerSerAsnTy-----Pr 602
Db      709 -----GAGACTACGCGACACCTCGCGGCCAGGAGGCGAGCTGATGATATACAC 761
Qy      602 ovalaITrpSerIleAspValLeuThrIleSerGlnGly----- 615
Db      762 CATCTGAGGCTGCGCATGTGCAATTCACACGAGAGGAGAACAGTGTGATCTC 821
Qy      616 -----AspTyThrIleuTrpTyraGpHeserIleAsnAsnGlnLysAs 630
Db      822 CAACCACTTGGTCTGCTGATACCTCCGCAAGCAACGCTACCATCAATGTTGCCCTTC 881
Qy      630 pGluTrpLysLysIleGlySerValSerValLysThrProGlnuTyThr----- 647
Db      882 C-----TTTACCAGAACCCCGCATGAGCTCCACCATCGCTGC 917
Qy      648 -----HisProLeuPheGluValaGlyH1 655
Db      918 CGGGGCCATGGCCCGCTGAGTGTGCTGCCGTGGGACACCCACCAAGATAGGCTG 977
Qy      655 sAsnGlnThrSerThrTyThrIleuAspMetAlaHISAsnArg-----ValLeuPr 672
Db      978 GCAGAAGGATGAGGACGAGCTTCCCGGCCCGCGAGAGAGACGATGACAGTGATGCC 1037
Qy      672 o-----AspPheThrLeuLysAsnLeuGlyLeuProPheAs 684
Db      1038 CGAGGATGACGCTTCTTCATTCGTCGAGAGATGAGAGACATCCGG----- 1086
Qy      684 nglYgIuLeuValValaPheArgGlnThrGlnInserSerSerGlySerLeuTrpAla-- 703
Db      1087 -----GTTTACAGCTGACAGCTCAGAACAGTGCAGAGCGCTCTGCAAA 1133
Qy      704 -AlaGlnGluThrVal-----HisI1 710
Db      1134 CGCAACCTGACTCTGTAGAGACGCCGCTATTTTTCGCGCCCTGTTAGACAGAACCGT 1193
Qy      710 elYsGlnGlyGluThrPheValTyryLysProValaGlnGlyProIlePro----- 727
Db      1194 CACCAAGGAGAAACACGCGCTCGACGTCATGCTGAGGAGGCCCGCCCTCGGCT 1253
Qy      728 -----AspGlySerTyraArgAlaThr-----LeuHISAlaPheValaAsnG1 741
Db      1254 GAATGACCAAGAAATACACACCCCTTTGGTGGTGACCGAGAGGCACTTTTTCAGACAGG 1313
Qy      741 yGlnGlnGlnLeuTyrLeu-----LysGlyLysArgAsnTyTrH 754
Db      1314 CAACCACTGCTGTATCATCGTGCAGCTGAGAGTACGAGGAGCGCGGAAA-----TACAC 1367

```

```

Qy      754 rValLysIleValaAsnGlyThrAlaValaIleGluSerSerGlnuIleArgVa 774
Db      1368 GTGTGAGATGTCATATACCTTGGCACCGAG-----AGAGGCAACGTTGCCCTGACTGT 1421
Qy      774 lPheProAsnProAlaArGAspTyValaGlnIleSerAlaProCysile 790
Db      1422 GATCCCTACGTCGACGCTGCGACCTCCCTCACATATACGTCGCCATCGCTG 1470

RESULT 11
AK018132
LOCUS
DEFINITION
Mus musculus adult male medulla oblongata cDNA, RIKEN full-length
enriched library, clone:6330406P08:homolog to KIAA1001 PROTEIN,
full insert sequence.
ACCESSION
AK018132
VERSION
AK018132.1 GI:12857709
KEYWORDS
HTC; CAP trapper.
SOURCE
Mus musculus (strain:C57BL/6J) adult male medulla oblongata cDNA to
mRNA, clone:lib:RIKEN full-length enriched mouse cDNA library
clone:6330406P08.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, Y., Nishi, K., Kitahara, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanuki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multichannel sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4
Kawai, J., Shigaawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Aizawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamataka, I.,
Saito, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Glisic, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsuno, Y., Nkaiado, I., Pesole, G.,
Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bulic, C.,
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamliya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzaletti, J., Mombaserts, P., Nordone, P.,
Ringo, B., Ringwald, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Sato, K., Schonbach, C., Seta, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-Oka, K., Wang, K. H., Weitz, C., Whitaker, C., Whiting, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S.
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)
JOURNAL
PUBMED
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

```


[illegible][illegible]


```

OY 558 npeasensr-----561
DB 441 ATTCATTTCTAAAGTATATTAATTAATTAACCAACCAAGCAAGAGCTTTTAAATCAA 352
OY 562 -----AspSerProaspGluIleArgThrProValAlaPheAlaIleuSerThrGlyAl 579
DB 381 ATTATATTAAGTAAAGTAAAGTAAATTAAG-----TTATCTACAGGAAC 340
OY 579 atrrAlaaspAspVal-----IleSerLeuGlyThrValMetAl 592
DB 339 ATACTATATACACTGTATATGCTTCACCAATTAACATTAATTAATGAAG-----ATGGAAG 283
OY 592 agluValProGlyGlySerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 611
DB 282 TAAACCCCAATTTCACTGATCTTATGATGACAGATTTGATGATTTACATCAATCAATCAC 223
OY 611 rleuSerGlu-----GlyAspThrThr 619
DB 222 CCTGATACAGTAAATTAATTAACAGACCTTCATCAATTCATCATGATGATGATGATGATG 163
OY 619 utrPlyrArgPheSerIleAsnAsnGlnIleAspGluThrPlyrIleGlySerValSe 639
DB 162 ATTCACCCCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 104
OY 639 rValIysThrProThrGluTyr 646
DB 103 CTTCACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 82

```

```

RESULT 15
LOCUS B1951857 944 bp mRNA linear EST 19-Oct-2001
DEFINITION HVSMEM0003E22f Hordeum vulgare green seedling EST library
HYCDNA0014 (Blumeria infected) Hordeum vulgare cDNA clone
B1951857
B1951857 GI:16295148

```

```

ACCESSION B1951857
VERSION B1951857
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare.

```

```

REFERENCE
AUTHORS Fritsch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,
J., Oates,R. and Main,D.
Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D.,
Development of a genetically and physically anchored EST resource
for barley genomics: Blumeria infected Morex (compatible) seedling
cDNA library
Unpublished (2001)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 425
Seq primer: AATTAACTCCTCACTAAAGG
High quality sequence start: 25
High quality sequence stop: 656.

```

```

TITLE
JOURNAL
COMMENT
FEATURES
source
1..944
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEM0003E22f"
/hyCDNA0014 (Blumeria infected)
/issue-type="green seedling leaf"
/lab_host="TUC121"
/Note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:

```

XhoI; Morex (mla) plants were greenhouse grown in the R
Wise lab at Iowa State University, Ames, IA; 7 day old
green seedlings were infected with isolate 5874 of
Blumeria graminis f. sp. hordei, and leaves were harvested
24, 48 and 72 hr post-inoculation and snap frozen (Wise).
In the rV Close lab at the University of California,
Riverside, total RNA was prepared from each sample pool,
equal quantities of all three RNA pools were combined,
poly(A) RNA was purified from the mixture, one primary
unamplified cDNA library was made, and 1 million p1u were
in vivo excised to give pBluescript SK(-) cDNA phagemids
(Chin). Phagemids were plated and picked at the Clemson
University Genomics Institute (CUGI) (Begum, Palmer,
Fritsch, Atkins and Wing). Plasmid DNA preparations, DNA
sequencing and sequence analysis were performed at CUGI
(Wing, Yu, Fritsch, Henry, Simmons, Oates, Rambo, Main).
The sequence has been trimmed to remove vector sequence
and contains a minimum of 100 bases of phred value 20 or
above. For more details on library preparation and
sequence analysis see
<http://www.genome.clemson.edu/projects/barley>. To order
this clone see <http://www.genome.clemson.edu/orders> Also
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley genomics. Barley Genetics Newsletter 31:29-30.
(<http://wheat.pw.usda.gov/gppages/bgn/31/cover.html>)"

```

BASE COUNT 200 a 268 c 306 g 170 t
ORIGIN

```

Alignment Scores:

```

Pred. No.: 3.37 Length: 944
Score: 99.00 Matches: 58
Percent Similarity: 32.80% Conservative: 24
Best Local Similarity: 23.20% Mismatches: 90
Query Match: 2.23% Indels: 78
DB: Gaps: 12

```

```

US-10-030-330-1 (1-843) x B1951857 (1-944)

```

```

OY 25 ThrLysGluArgAlaLeuSerLeuAlaArgLeuAlaLeuArgGlnValSerLeuArgMet 44
DB 182 ACTAGCCGTCACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 241
OY 45 GlyGlnThrAlaValSerAspIleSerIleAspThrValTyrArgGlnIleAspAla 64
DB 242 GCCGCTACCCCGCTGTCGAC-----CGCAAGGCTCCGCA 277
OY 65 GluArgGlyIleThrSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 83
DB 278 GAGCCTTCCGAGAAACAGAGGACGAGCGCGGACGACGACGACGACGACGACGACGACGAC 337
OY 83 -----83
DB 338 AAGTTCATGCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 397
OY 84 -----ArgGlyAsnAsnGlnGlyTyrAlaLeuValAlaAlaAsp-----Asp 97
DB 398 TACGAGAGCGCCAGCAAGAGGACGACGACGACGACGACGACGACGACGACGACGACGAC 457
OY 98 ArgIleProThrIle-----LeuAlaTyrSerProIleGlyArgPheAspMetLasp 114
DB 458 AGGTGCCCAAGTCCGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 517
OY 115 SerMetProAspAsnLeuArgMetTyrPleuGlnIleTyrAspGlnIleGlyLeuIle 134
DB 518 -----CGCAACTGATCATGATCATGATCATGATCATGATCATGATCATGATCATGAT 559
OY 135 LeuSerGlyLys-----AlaGlnLeuAsnGlnGlnIleLeuArgThrGlnGlyValPro 152
DB 560 TTAGCGGACCAATTCATGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 613
OY 153 AlaGluValHisAlaLeuMetAspAsnGlnIleHisPheAlaAsnAspProMetArgTyrAsn 172
|||||:

```

DB 614 GCAGACTGAAA-----TTGATACCTGACAAATGACA 649

QY 173 GINGLYTYPRTPRAsnAsnLysGluProLeuLeuProAsnGlyAsnHisAlaTyrThr 192

DB 650 GCTCCATAC-----CCAGATACGACTCGATGGG 682

QY 193 GTCYsValAlaThrAlaAlaGlnLeuMetArgTyrHis-----SerTyrPro 209

DB 683 GGAGGCGCTTCTACTCCAGAGCCAGCAGCGGTTAGATAGGCTCAGTCGCCG 742

QY 210 LeuGlnGlyGluGlySerPheAspTyrHisAlaGlySerLeuValGlyAsnTyrSerGly 229

DB 743 -----CGGCTAGGATTTCTCTCGAAGAGATGATGG 775

QY 230 ThrPheGlyGluMetTyrAsp---Triple 238

DB 776 AAGTTGGGTTATGCTATGGGATGGGTC 805

CHUT 16
5828

DESCRIPTION BM815828 765 bp mRNA linear EST 05-MAR-2002
EST539322 BNR Medicago truncatula/Meloidogyne incognita mixed EST
library cDNA clone PBNIR-28F2, mRNA sequence.
BM815828
BM815828.1 GI:19151842

ACCESSION
VERSION
KEYWORDS
SOURCE
ORIGIN
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Medicago truncatula/Meloidogyne incognita mixed EST library.
Medicago truncatula/Meloidogyne incognita mixed EST library.
Eukaryota: mixed EST libraries.
1 (bases 1 to 765)
Cheung, F., Tsai, J. and Fraser, C.M.
ESTs from roots of Medicago truncatula after infection with the
nematode Meloidogyne incognita
Unpublished (2002)
Contact: Bird, DM
Plant Nematode Genetics Group
North Carolina State University
Box 7616, Raleigh, NC 27695, USA
Tel: 919 515 6813
Fax: 919 515 9500
Email: david_bird@ncsu.edu
TIGR sequence name: MTOBJ25K More information is available at:
www.medicago.org
Seq primer: SKmod (CTA GAA CTA gta gat CC).
Location/Qualifiers
1. 765
/organism="Medicago truncatula/Meloidogyne incognita mixed
EST library"
/cultivar="A17"
/db_xref="taxon:188702"
/clone="PBNIR-28F2"
/clone_11b="BNIR"
/tissue_type="3 week old roots"
/dev_stage="3 days after infection with Meloidogyne
incognita second stage larvae"
/lab_host="XtDR"
/note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the XhoI/XbaI vector from
stratagene and packaged using GigaPack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-zap phage using Ex-assist
helper phage and propagated in XtDR cells."

BASE COUNT 217 a 174 c 145 g 229 t

ORIGIN

Alignment Scores:
Pred. No.: 2.59 Length: 765
Score: 98.50 Matches: 54
Percent Similarity: 31.87% Conservative: 26
Best Local Similarity: 21.51% Mismatches: 86
Query Match: 2.22% Indels: 85

DB: 14 Gaps: 9

US-10-030-330-1 (1-843) x BM815828 (1-765)

QY 395 AlAlaYsThrProAlaGlnLeuThrAspAlaLeuProLeuAlaLeuLysAspIle 414

DB 193 GCAGGCTATGATTCACGATTTGGCTTGAACCCAGCCCTAAATTTAGGACTTCCGCTCA 252

QY 415 GluAlaGluTyrLysSerGluSerGlyLeuAsnValGly----- 427

DB 253 ATTGAAGCAAGATATATCTATGGTGTGTAAGTTGGATTCGATACAACTGCTTCA 312

QY 428 TYSerIleTyrAsnThrGlyGluGlnSerAsnLeuAspLeuGlyTyrArgLeuAsn 447

DB 313 TCTCTCAACATATTAATGCCGGATTCCTTCATTAACCAAGATTTCTCGAACAACATTATG 372

QY 448 LysAlaAspGlyGluValIleGluValLysThrSerIleAsnIleSerTyrTyrGly 467

DB 373 CTGGCTGAT-----AAAGACAGTCTCTGAAGGATCTTACATTCAT 414

QY 468 TYRGLYGLuHisProGluSerPheSerLeuAla-----ProAsnGlnLeuSerGln 484

DB 415 TATGTCGACCGCCGATGATGATTAACAGTCTCTCGAAGCTGCTCAAGTTGTCTCT 474

QY 485 GlyIleAsnThrIleThrLeuLeuTyrArgTyrGlyThrGluGlnTyrProVal 504

DB 475 TCTGAGAACACATTTACTCTT-----GGAGCTCAG----- 507

QY 505 ArgHisAlaGlnGlyTyrValAsnSerIleValAsnThrThrAspProAsnAsn 524

DB 508 -----TCAATGATGCCAANAACA 525

QY 525 ValValAlaThrValAspAsnAsnGluGlyLysLeuSerIleVal----- 539

DB 526 GTTTGAAGACTCGATTCACATGATGATGAGGCAAGGCTCTCCAAATGCCACAGCGGTGG 585

QY 540 ---ProAsnSerPheValAlaAspLeuAsnSerTyrGluHisSerThrIleThrValGln 558

DB 586 AGGCCAAATTCATC-----ATAACCTATCT 612

QY 559 PheAsnSerAspSerProAspGluIleArgThrProValAlaPheAlaLeuSerThrGly 578

DB 613 GCCGAGTATGATGCCAANAATCATTTGTCACCTGCCAANAATGCGTCTCTCT 669

QY 579 AlaThrAlaAspAspValIleSerLeuGlyTyrValMetAlaGluValProGlyLysSer 598

DB 670 -----TCT 672

QY 599 SerAsnTyrProValAlaTyrSerLysAsp-----Val 609

DB 673 CTCAGGCTTAACCTGGTTTGGCAATCAATCTTATTCACAGCTTTTAACCTTACCAATT 732

QY 610 LeuThrLeuSerGluGlyAspTyrThrLeuTyr 620

DB 733 CTATCATTTTACATTTGTCGCAATTCACCTCTG 765

RESULT 17
BF383380/c 1378 bp mRNA linear EST 27-NOV-2000
LOCUS 602045042P1 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:4194395 5',
DEFINITION mRNA sequence.
BF383380
ACCESSION BF383380.1 GI:11364685
VERSION
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1378)
NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

OY		594	ValProGlygylserSerAsnTyrProValAlaValrpsLysaspValLeuThrLeuSer	613
Dd		467	GAGACGGTGCGT-----CCGTTCCTGCCTGCCTATGCATGCTTTGGCGAAGA	514
OY		614	GlUglyAspTyrThrLeuTrpTyrArgPhe---SerIleasnAnsnLnLysaspGluTrp	632
Dd		515	AACGGTGCCTTTGTTGGGTACCACACTGAGATCTCTTAGCAGCATGAAGATTCTGG	574
OY		633	LysLysIleGlySerVal-----SerValLysThrProThrGluLysThrHisProLeu	650
Dd		575	GAGGTTTGTCGAGAGAGATTGATGTCGAGAGAGTTCCACACGACATTACTCA----	631
OY		651	PheGluValAlaGlyHisAsnGlnThrSerThrTyr	661
Dd		632	---GAGTACGCCCTATGAGGAGACTGCAGCTTTAT	661
RESULT 19				
CNS077A/c				
LOCUS		1037 bp	DNA	linear GSS 08-JUL-2001
DEFINITION		J3 end of clone BBOAA010C12 of library BBOAA from strain CBS 4732		
ACCESSION		AF432436		
VERSION		AF432436.1		
KEYWORDS		GT:12215850		
SOURCE		GSS.		
ORGANISM		<i>Pichia angusta</i> .		
REFERENCE		<i>Pichia angusta</i>		
AUTHORS		Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia. 1 (bases 1 to 1037)		
TITLE		Souci�t,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolo�lin-Fukuhara,M., Bon,E., Brotier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Leppigne,A., Llorente,B., Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potter,S., Saurin,M., Tekala,F., Toïfano-Nioche,C., Wesolowski-Louvel,M., Winker,P. and Weissenbach,J. Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies FEBS Lett. 487 (1), 3-12 (2000) 20584711		
JOURNAL MEDLINE PUBMED		2 (bases 1 to 1037)		
REFERENCE AUTHORS		Blandin,G., Llorente,B., Malpertuy,A., Winker,P., Artiguenave,F. and Dujon,B. Genomic exploration of the hemiascomycetous yeasts: 13. <i>Pichia</i> <i>angusta</i> FEBS Lett. 487 (1), 76-81 (2000) 20584723		
JOURNAL MEDLINE PUBMED		3 (bases 1 to 1037)		
REFERENCE AUTHORS		Genoscope. Direct Submission Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremlieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqlife@genoscope.cns.fr Web : www.genoscope.cns.fr) This GSS is part of a random genomic sequencing program of thirteen yeast species: <i>Saccharomyces bayanus</i> var. <i>uvarum</i> , <i>Saccharomyces</i> <i>exiguus</i> , <i>Saccharomyces servazii</i> , <i>Zygosaccharomyces rouxii</i> , <i>Saccharomyces kluyveri</i> , <i>Kluyveromyces thermotolerans</i> , <i>Kluyveromyces</i> <i>lactis</i> var. <i>lactis</i> , <i>Kluyveromyces marxiianus</i> var. <i>marxiianus</i> , <i>Pichia</i> <i>Candida tropicalis</i> and <i>Xarowia lipolytica</i> . Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert. Location/Qualifiers 1..1037		
COMMENT				
FEATURES				
source				

msc-feature <138. >422
 /note="similar to Saccharomyces cerevisiae ORF YGL191w [COX13 ; cytochrome c oxidase chain via]"
 /evidence-not_experimental
 msc-feature
 /note="similar to Saccharomyces cerevisiae ORF YGL190c [CDC55 ; ser/thr phosphatase 2A regulatory subunit B]
 1 putative frameshift(s)"
 /evidence-not_experimental
 BASE COUNT 306 a 219 c 206 g 305 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 5.27 Length: 1037
 Score: 98.00 Matches: 47
 Percent Similarity: 34.91% Conservative: 27
 Best Local Similarity: 22.17% Mismatches: 57
 Query Match: 2.21% Indels: 81
 Gaps: 10

0-030-330-1 (1-843) x CNS077a6 (1-1037)

QY 289 ArgAsnAspNheArgTyrLysArg-----SerLeuGlnLeuHis---Val 302
 |||||
 Db 993 AGAATAAAGTCTCTCTATATGATGATGCAATTCAGCTCAACGCGAGATACATGTGTC 934
 QY 303 ArgAlaLeuTyr-ThrSerGlnGluTrp----- 311
 |||||
 Db 933 GAGACATTAATATGACGTTAGATATGAGATATAGCGATGAGCTCGAACAATCAAAAC 874
 QY 312 -----HisAspMetIleArgGlyLeuLeuAlaSerGlyArgProValTyrTyrAl 328
 |||||
 Db 873 AATCAACATTCATGAGCAGCTCCGGATGGCTG----- 840
 QY 328 aclysnasnaglnserIleGlyHisAlaPheValCysaspGlyTyrAlaSeraspGly 348
 |||||
 Db 839 -----TGTGACACATATGAGATATGACGCGAT 814
 QY 348 rPhe--HisPheAsnTrpGlyTyrGlyValSeranglyPheTyrLysLeuThrIle 367
 |||||
 Db 813 CTTTGATTAATTTGAGCTGCAATTCAGCGAGACAAACAA----- 774
 QY 367 uLeuSerProThrSerLeuGlyIleGlyGlyGlnGlyIleGlyPheThrIleTyrGlnG 387
 |||||
 Db 773 -----TCCATATATGACGCTTCATACAAATTAATTCATGATCTACCAAAA 727
 QY 387 uIleIleThrGlyIleGlnProAlaIleThrProAlaGlnIleGlyThrAspAlaLeuP 407
 |||||
 Db 726 TGTGTGGGTAACTCAGTCCAGAAAGAG----- 696
 QY 407 oIleLeuAlaLeuLysAspIleGlnIleGlnIleTyrLysSerGlyLeuAsnValG 427
 |||||
 Db 695 -----ATTGTCTTCAAGCGGATAGTACACATTTAAAGCAAGAAA----- 654
 QY 427 yTyrSerIleTyrAsnThrGlyGlnGlnIleSerAsnLeuAspLeuGlyTyrArgLeu 447
 |||||
 Db 653 -----ATGAATTCACAACAGCGGAAGATCTAACAAATG-----ATGAA 616
 QY 447 nLysAlaAspGlyGlnValIleGlnValIleYsthrSerSerIleAsnIleSerTyrG 467
 |||||
 Db 615 TGAATGAGCCTTCGACACCATGACACTTCAAAAATTAATTTACACTGAGCTGG----- 561
 QY 467 yTyrGlyGlnHisProGlyLeuSerPheSerLeuAla 478
 |||||
 Db 560 -----CATCTCAGCAAAACTCAGTTCGC 537
 RESULT 20
 B1838400/c 1149 bp mRNA linear EST 04-OCT-2001
 LOCUS 603081361f1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5222511 5',
 DEFINITION mRNA sequence.
 ACCESSION B1838400
 VERSION B1838400.1 GI:15949950

KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1149)
 AUTHORS NIH-MGC: <http://mhc.ncl.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM1559 row: 3 column: 16
 High quality sequence start: 34
 High quality sequence stop: 110.
 Location/Qualifiers
 1. 1149
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5222511"
 /clone_1lb="NIH_MGC_120"
 /lab_host="DH10B"
 /note="Organ: pooled pancreas and spleen; Vector: pCMV-SPORT6; Site: 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of spleen and pancreas from 28 yo male. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1.2-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 025. Note: this is a NIH_MGC Library."

BASE COUNT 212 a 311 c 345 g 281 t
 ORIGIN

Alignment Scores:
 Pred. No.: 6.41 Length: 1149
 Score: 98.00 Matches: 34
 Percent Similarity: 46.15% Conservative: 20
 Best Local Similarity: 29.06% Mismatches: 33
 Query Match: 2.21% Indels: 30
 Gaps: 9

US-10-030-330-1 (1-843) x B1838400 (1-1149)

QY 147 ArgTThrGlnGlyValProAlaGlnValHisAlaLeuMetaspn-----Gly 162
 |||||
 Db 294 CGAGCTAGCGGCGCTTCCACGGAT---CACACCCTGGTTCACACGCGGCTCAGTCGA 238
 QY 163 HisPheAlaAsnAspProMetArgTyrPasnGlnGlyTyrProThrPasnAsnLysGluPro 182
 |||||
 Db 237 CCAGAGACACATGATCATGACCGCTGACCAACAGGACATCT-----GCT 193
 QY 183 LeuLeuProAsnGlyAsnHisAlaTyrThrGlyCysValAlaThrAlaAlaGlnIle 202
 |||||
 Db 192 CTGATGCTTTCGCAAGTACGCGGTGAGGAGGACACCAATGACAGTACAGAGCAC-- 136
 QY 203 MetArgTyrHisSerTrpProLeuGlnGlyGlnGlySerPheAspTyrHisAlaGlySer 222
 |||||
 Db 135 ---TGATTCACATTTATCT-----CACAGGGGGCA 106
 QY 223 LeuValGlyAsnTrpSerGlyThrPheGlyGlnMet-----TyrAspTrpIleAsn 239
 |||||
 Db 105 AATGTGTGATCATTTGGCGCCCTCTTG---GACCTACTCTCCCTTCAGCTGG----- 55
 QY 240 MetProGlyAsnProAspLeuAspAsnLeuThrGlnSerGlnValAspAla 256
 |||||
 Db 54 GCCCCAGGAGGTGC-----CACACACCAACAGCTGATCATGATCA 13

Mon Jun 2 11:44:00 2003

us-10-030-330-1.p2n.rst

Page 22

Search completed: June 2, 2003, 01:02:24
Job time : 2993 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 13:21:41 ; Search time 42 Seconds
(without alignments)
2674.530 Million cell updates/sec

Title: US-10-030-330-1

Perfect score: 4438
Sequence: 1 MKSFLAIIVMLFGIIMOGH.....PNGAVILKVDGTTTINIVH 843

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: A.Geneseq_101002.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4438	100.0	843	20	AAV34485	Porphyromonas ging
2	4438	100.0	849	21	AAV34360	Porphyromonas ging
3	3946.5	88.9	844	21	AAV36100	Porphyromonas ging
4	1078	24.3	840	21	AAV36099	Porphyromonas ging
5	494	11.1	398	21	AAV81812	Porphyromonas ging
6	493	11.1	398	23	ABP29579	S. pyogenes cystei
7	489	11.0	398	17	AAW07898	Streptococcus poly
8	489	11.0	398	17	AAW07898	Streptococcus poly
9	449	10.1	398	21	AAV36098	S. pyogenes speb g
10	190	4.3	540	20	AAV34573	Streptococcus pyog
						Porphyromonas ging

11	190	4.3	563	20	AAV34430	Porphyromonas ging
12	158.5	3.6	921	20	AAV34522	Porphyromonas ging
13	158.5	3.6	922	20	AAV34521	Porphyromonas ging
14	158.5	3.6	925	20	AAV34520	Porphyromonas ging
15	158.5	3.6	938	20	AAV34392	Porphyromonas ging
16	154	3.5	1325	22	AAV98256	Escherichia coli p
17	153	3.4	2652	22	ABG11747	Novel human diagno
18	148.5	3.3	821	20	AAV34479	Porphyromonas ging
19	148.5	3.3	869	20	AAV34354	Porphyromonas ging
20	148.5	3.3	1352	22	AAV63962	Amino acid sequenc
21	147.5	3.3	699	20	AAV08471	F. balustium CP70
22	142.5	3.2	390	20	AAV34444	Porphyromonas ging
23	142.5	3.2	391	20	AAV34432	Porphyromonas ging
24	142.5	3.2	940	23	ABV47334	Porphyromonas ging
25	140.5	3.2	647	16	AAV11978	Listeria monocytog
26	140	3.2	934	18	AAV34451	Mycothacterium gall
27	140	3.2	934	18	AAV31367	Enterohaemorrhagic
28	140	3.2	1349	23	ABV47323	Listeria monocytog
29	140	3.2	10182	23	ABP38314	Staphylococcus epi
30	139.5	3.1	814	10	AAV90681	Kex2 endoprotease.
31	139.5	3.1	814	18	AAV25685	Kex2 endoprotease.
32	139.5	3.1	1256	13	AAV27746	Muramidase release
33	139.5	3.1	2657	22	ABG04691	Novel human diagno
34	139	3.1	1752	23	ABV05030	Bifidobacterium bl
35	138.5	3.1	648	15	AAV56973	PMGA 1.2 protein o
36	138.5	3.1	2902	22	AAV46351	H. pylori HPN165 p
37	138	3.1	1398	17	AAV87008	Protease. Pyrococ
38	138	3.1	1398	18	AAV24124	Pyrococcus furiosu
39	138	3.1	1398	20	AAV94839	W0956926 Seq ID 6
40	137	3.1	539	21	AAV54011	Amino acid sequenc
41	134.5	3.0	1648	23	ABV54925	Lactococcus lactis
42	134	3.0	1074	22	ABV24648	Novel human diagno
43	133.5	3.0	1310	23	ABV48256	Listeria monocytog
44	133	3.0	704	20	AAV11683	Sulfolobus fucos-co
45	133	3.0	1376	22	ABV52592	Escherichia coli p

ALIGNMENTS

RESULT 1	
AAV34485	
ID	AAV34485 standard; Protein; 843 AA.
XX	
AC	AAV34485;
XX	
DT	25-AUG-1999 (first entry)
XX	
DE	Porphyromonas gingivalis protein PG28.
XX	
XX	Porphyromonas gingivalis; Pg. periodontal disease; gingivitis;
KW	vacuole; antigenic.
XX	
OS	Porphyromonas gingivalis.
XX	
PN	W0929870-AL.
XX	
PD	17-JUN-1999.
XX	
PF	10-DEC-1998; 98WO-AU01023.
XX	
PR	04-AUG-1998; 98AU-0005028.
PR	10-DEC-1997; 97AU-0000839.
PR	31-DEC-1997; 97AU-0001182.
PR	30-JAN-1998; 98AU-0001546.
PR	10-MAR-1998; 98AU-0002264.
PR	09-APR-1998; 98AU-0002911.
PR	23-APR-1998; 98AU-0003128.
PR	05-MAY-1998; 98AU-0003338.
PR	22-MAY-1998; 98AU-0003654.
PR	29-JUL-1998; 98AU-0004917.
XX	
PA	(CSLC-) CSL LTD.

XX Agius CT, Barr IG, Hocking DM, Margets MB, Patterson MA;
 PI-Ross BC, Rothel LJ, Webb EA;
 XX WPI: 1999-385613/32.
 DR N-PSDB: AAX91703.
 XX
 PT Antigenic Porphyromonas gingivalis peptides for preventing
 gingivitis
 XX
 PS Claim 1; Page 466-468; 568pp; English.
 CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (Pg) polypeptide sequences given in AAY34318 to
 CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the Pg polypeptides. The Pg polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The Pg polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.
 SU Sequence 843 AA:
 Query Match 100.0%; Score 4438; DB 20; Length 843;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKSFLAIYMLGIMOGSHAPVTERALSLRLALROVSLMGOTAVSDKTSIDYVR 60
 DB 1 MKSFLAIYMLGIMOGSHAPVTERALSLRLALROVSLMGOTAVSDKTSIDYVR 60
 QY 61 OGDAERGITSQEGSPAYFYVANRGNNEGATVAADRIPTILAYSPIGRFMDMPDNL 120
 DB 61 OGDAERGITSQEGSPAYFYVANRGNNEGATVAADRIPTILAYSPIGRFMDMPDNL 120
 QY 121 RMWLOIYDOETIGLISGKAOLEIIRTEGVPALVHALMDNGHFPANDPMKNOGYPMNNK 180
 DB 121 RMWLOIYDOETIGLISGKAOLEIIRTEGVPALVHALMDNGHFPANDPMKNOGYPMNNK 180
 QY 181 EPLLPGNNAHYTCVATTAQIMRYHSMPLQEGSGFDYHAGSLVGNMSEFGEMYPWINM 240
 DB 181 EPLLPGNNAHYTCVATTAQIMRYHSMPLQEGSGFDYHAGSLVGNMSEFGEMYPWINM 240
 QY 241 PCNPDLNLTQSGVDAYATLMRDVSAVSMSFYENGSGTYVYVGLARNNEFYKRSLOL 300
 DB 241 PCNPDLNLTQSGVDAYATLMRDVSAVSMSFYENGSGTYVYVGLARNNEFYKRSLOL 300
 QY 301 HVRLATTSQEMHDMIRGELASGRPVYVAGNNSIGHAFVCDGASDGFHFMNGGCVSN 360
 DB 301 HVRLATTSQEMHDMIRGELASGRPVYVAGNNSIGHAFVCDGASDGFHFMNGGCVSN 360
 QY 361 GFYKLTLLSPTSLGIGEGIGFTIYOETIIRTEGVPALVHALMDNGHFPANDPMKNOGYPMNNK 420
 DB 361 GFYKLTLLSPTSLGIGEGIGFTIYOETIIRTEGVPALVHALMDNGHFPANDPMKNOGYPMNNK 420
 QY 421 ESGLANVGYSTYNTGEEQSNLDLGYRLNKADGEVIEVKTSSINISWYGYGHPSPSLAPN 480
 DB 421 ESGLANVGYSTYNTGEEQSNLDLGYRLNKADGEVIEVKTSSINISWYGYGHPSPSLAPN 480
 QY 481 QLSQGINITITLLYRRRTGTEOMEVPRHAOGGYVNSIVNTTDPNNVVYVDNNEGKSIYP 540
 DB 481 QLSQGINITITLLYRRRTGTEOMEVPRHAOGGYVNSIVNTTDPNNVVYVDNNEGKSIYP 540
 QY 541 NSFVADLNSYESTITVOFNSDSPDEIRTPVAALSTGATADVISLGMVMAVPGSSN 600
 DB 541 NSFVADLNSYESTITVOFNSDSPDEIRTPVAALSTGATADVISLGMVMAVPGSSN 600
 QY 601 YPVVMSKDVLTSEGGYTLMYRFSINNOKDEMCKISVSYPKPTETHTPLFEVGHNOTST 660
 DB 601 YPVVMSKDVLTSEGGYTLMYRFSINNOKDEMCKISVSYPKPTETHTPLFEVGHNOTST 660
 QY 661 YTLDMANRVLDPFTLKNLGLPENGELVYVFFROTQSSGSLMAAQTVAHKQGETVYKP 720

DB 661 YTLDMANRVLDPFTLKNLGLPENGELVYVFFROTQSSGSLMAAQTVAHKQGETVYKP 720
 QY 721 VVEGPIPDGSYRATILAFVNGOOLYLKGRKNTVAVTNGTAVALESSSEIRVFPNPAR 780
 DB 721 VVEGPIPDGSYRATILAFVNGOOLYLKGRKNTVAVTNGTAVALESSSEIRVFPNPAR 780
 QY 781 DVEISAPCIPQETSIIIFDLSGRIYMKNSLSAGHRMDVSRPLPNCATILKYDGYTTKIN 840
 DB 781 DVEISAPCIPQETSIIIFDLSGRIYMKNSLSAGHRMDVSRPLPNCATILKYDGYTTKIN 840
 QY 841 IVH 843
 DB 841 IVH 843
 RESULT 2
 AAY34360
 ID AAY34360 standard; Protein; 849 AA.
 AC AAY34360;
 XX
 DF 25-AUG-1999 (first entry)
 DE Porphyromonas gingivalis protein PG28.
 XX Porphyromonas gingivalis; Pg; periodontal disease; gingivitis;
 KM vaccine; antigenic.
 XX Porphyromonas gingivalis.
 OS MO929870-A1.
 PN
 XX
 PD 17-JUN-1999.
 XX
 PF 10-DEC-1998; 98WC-AU01023.
 XX
 PR 04-AUG-1998; 98AU-0005028.
 PR 10-DEC-1997; 97AU-0000839.
 PR 31-DEC-1997; 97AU-0001182.
 PR 30-JAN-1998; 98AU-0001546.
 PR 10-MAR-1998; 98AU-0002264.
 PR 09-APR-1998; 98AU-0002311.
 PR 23-APR-1998; 98AU-0003128.
 PR 05-MAY-1998; 98AU-0003338.
 PR 22-MAY-1998; 98AU-0003654.
 PR 29-JUL-1998; 98AU-0004917.
 XX
 PA (CSLC-) CSL LTD.
 XX
 PI Agius CT, Barr IG, Hocking DM, Margets MB, Patterson MA;
 PI-Ross BC, Rothel LJ, Webb EA;
 XX WPI: 1999-385613/32.
 DR N-PSDB: AAX91578.
 XX
 PT Antigenic Porphyromonas gingivalis peptides for preventing
 gingivitis
 XX
 PS Claim 1; Page 323-324; 568pp; English.
 CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (Pg) polypeptide sequences given in AAY34318 to
 CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the Pg polypeptides. The Pg polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The Pg polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.
 SU Sequence 849 AA:

Query Match 100.0%; Score 4438; DB 20; Length 849;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 843; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKSFLAIYMLFGIMOGHSAFYTKERALSRLALROYSLRMGQTAVSDDKISIDVYR 60
 DB 7 MKSFLAIYMLFGIMOGHSAFYTKERALSRLALROYSLRMGQTAVSDDKISIDVYR 66
 QY 61 QGDAERGITSQEGSPAYFYVANGNNEGIALVAADRIPITLAYSPIGRFDMDSMDNL 120
 DB 67 QGDAERGITSQEGSPAYFYVANGNNEGIALVAADRIPITLAYSPIGRFDMDSMDNL 126
 QY 121 RMWLOIYDOEIGLISGKAQNLNEILRTGEPVAVHLMONGHAFANDPMRNOGYPMNNK 180
 DB 127 RMWLOIYDOEIGLISGKAQNLNEILRTGEPVAVHLMONGHAFANDPMRNOGYPMNNK 186
 QY 181 EPLLPGNHAHYTCGVAATAAOLMRYSWPLQGGSEFDYHAGSLVGNMSTFGEMYDWINM 240
 DB 187 EPLLPGNHAHYTCGVAATAAOLMRYSWPLQGGSEFDYHAGSLVGNMSTFGEMYDWINM 246
 QY 241 PGMPDIDNLTQSOVDVYATLMRDVSAVSMSFEYENGSGTYSYVVGALRNFRKRSLOL 300
 DB 247 PGMPDIDNLTQSOVDVYATLMRDVSAVSMSFEYENGSGTYSYVVGALRNFRKRSLOL 306
 QY 301 HVRALYTSQEMHDMIRGELASGRPYVYAGNNOISGHAFCVDCGASDGFHFNNMGWGSVN 360
 DB 307 HVRALYTSQEMHDMIRGELASGRPYVYAGNNOISGHAFCVDCGASDGFHFNNMGWGSVN 366
 QY 361 GFYKTLTSPISIGIGEGIGFTIYOEITIGIEPANTPAAGADALPILAKDIEAEYKS 420
 DB 367 GFYKTLTSPISIGIGEGIGFTIYOEITIGIEPANTPAAGADALPILAKDIEAEYKS 426
 QY 421 ESSLANGSYIYNNGEESQNDLGYRLNKADGEVIEVKTSSINISWYGGHSPESFLAPN 480
 DB 427 ESSLANGSYIYNNGEESQNDLGYRLNKADGEVIEVKTSSINISWYGGHSPESFLAPN 486
 QY 481 QLSQGINITITLLYRRTGTEQMEPVRAHQAQGVNSIKVNTDPNNVVYTVNNEGKLSIVP 540
 DB 487 QLSQGINITITLLYRRTGTEQMEPVRAHQAQGVNSIKVNTDPNNVVYTVNNEGKLSIVP 546
 QY 541 NSFVADLNSTYESTITVQENSQSPDEIRFPVAFALSTGATADYISLGWMAEVPGGSSN 600
 DB 547 NSFVADLNSTYESTITVQENSQSPDEIRFPVAFALSTGATADYISLGWMAEVPGGSSN 606
 QY 601 YPVMYKDVLTLSSEGYTLMYRFSINNOKDEMKKISVSQVKTPTETHRPFYVGHQNST 660
 DB 607 YPVMYKDVLTLSSEGYTLMYRFSINNOKDEMKKISVSQVKTPTETHRPFYVGHQNST 666
 QY 661 YTLDMANRYLPDFTLKNLGLPENGELVYVFPOTQSSSGSLMAAOETVHIKOGTFYKRP 720
 DB 667 YTLDMANRYLPDFTLKNLGLPENGELVYVFPOTQSSSGSLMAAOETVHIKOGTFYKRP 726
 QY 721 VVEGPIPDGSYRATLHAFVNGOOLYLKGRNRYTVKIVNGTAVAEIESTEIRVFPNPAR 780
 DB 727 VVEGPIPDGSYRATLHAFVNGOOLYLKGRNRYTVKIVNGTAVAEIESTEIRVFPNPAR 786
 QY 781 DVEIETAPCPOTESTIITFDLSKTIYMKNSLSAGHGMDSRLPNAGYILKAVGYTTKIN 840
 DB 787 DVEIETAPCPOTESTIITFDLSKTIYMKNSLSAGHGMDSRLPNAGYILKAVGYTTKIN 846
 QY 841 IVH 843
 DB 847 IVH 849

RESULT 3
 AAB36100 standard; Protein; 844 AA.
 AC AAB36100;
 XX
 XX 16-FEB-2001 (first entry)
 XX

DE Porphyromonas gingivalis periodontain.
 XX Porphyromonas gingivalis; periodontain; antiinflammatory; antibacterial;
 KW amidolytic; alpha-1-proteinase inhibitor; periodontitis; gingivitis.
 XX Porphyromonas gingivalis.
 XX W0200063394-A2.
 XX 26-OCT-2000.
 XX 20-APR-2000; 2000MO-US10574.
 XX 21-APR-1999; 99US-0130436.
 XX (UYGE-) UNIV GEORGIA RES FOUND INC.
 XX (TRAV/) TRAVIS J.
 XX (POTE/) POTEPA J.
 XX (NELS/) NELSON D.
 PI Travis J, Potempa J, Nelson D;
 XX WPI; 2000-679600/66.
 XX Novel oral bacterial periodontain polypeptide for treating periodontal
 PT diseases, has amidolytic activity for cleavage of non-denatured human
 PT alpha1-proteinase inhibitor at reactive site loop region of inhibitor
 PT
 XX
 XX Example 1; Fig 1; 55pp; English.
 CC The present sequence is given in a specification relating to novel
 CC oral bacterial polypeptide referred to as periodontain. The polypeptide
 CC has amidolytic activity for cleavage of denatured polypeptides and
 CC non-denatured serpin polypeptides. It has amidolytic activity for
 CC cleavage of a non-denatured human alpha-1-proteinase inhibitor at a
 CC reactive site loop region of the inhibitor. Periodontain is useful for
 CC inhibiting the peptidase activity and reducing periodontitis, loss of
 CC tooth attachment and periodontal pocket formation, and for reducing
 CC growth of bacteria, preferably P. gingivalis in vitro or in vivo.
 CC It is useful for protecting an animal from a disease caused by
 CC P. gingivalis and for treating periodontal diseases, including
 CC gingivitis and periodontitis.
 CC
 SQ Sequence 844 AA;
 Query Match 88.9%; Score 3946.5; DB 21; Length 844;
 Best Local Similarity 89.8%; Pred. No. 4; 66-289;
 Matches 758; Conservative 4; Mismatches 81; Indels 1; Gaps 1;

QY 1 MKSFLAIYMLFGIMOGHSAFYTKERALSRLALROYSLRMGQTAVSDDKISIDVYR 60
 DB 1 MKSFLAIYMLFGIMOGHSAFYTKERALSRLALROYSLRMGQTAVSDDKISIDVYR 60
 QY 61 QGDAERGITSQEGSPAYFYVANGNNEGIALVAADRIPITLAYSPIGRFDMDSMDNL 120
 DB 61 QGDAERGITSQEGSPAYFYVANGNNEGIALVAADRIPITLAYSPIGRFDMDSMDNL 120
 QY 121 RMWLOIYDOEIGLISGKAQNLNEILRTGEPVAVHLMONGHAFANDPMRNOGYPMNNK 180
 DB 121 RMWLOIYDOEIGLISGKAQNLNEILRTGEPVAVHLMONGHAFANDPMRNOGYPMNNK 180
 QY 181 EPLLPGNHAHYTCGVAATAAOLMRYSWPLQGGSEFDYHAGSLVGNMSTFGEMYDWINM 239
 DB 181 EPLLPGNHAHYTCGVAATAAOLMRYSWPLQGGSEFDYHAGSLVGNMSTFGEMYDWINM 240
 QY 240 PGMPDIDNLTQSOVDVYATLMRDVSAVSMSFEYENGSGTYSYVVGALRNFRKRSLOL 299
 DB 241 PGMPDIDNLTQSOVDVYATLMRDVSAVSMSFEYENGSGTYSYVVGALRNFRKRSLOL 300
 QY 300 LHVRLYTSQEMHDMIRGELASGRPYVYAGNNOISGHAFCVDCGASDGFHFNNMGWGSVN 359
 DB 301 LHVRLYTSQEMHDMIRGELASGRPYVYAGNNOISGHAFCVDCGASDGFHFNNMGWGSVN 360

Db 689 FQAKREHXSQACETKVLSPEXTANSLSYTNALFPDGRY-----YIYAREGGFMDPXD 742
 QY 747 LKGRNRYTKYNGTAVEAIESE--EIRYFPNPAOYVEISAPCIPOETSIILFDLSCK 804
 Db 743 LFGDYRRXXXTTDLSSDIAGCDVSTIYXXPNPAHDYVXAIPTVAGXXKXFLDIQGR 802
 QY 805 IYKKNLSLGHGMDVSRLLPNCAYILKYGGYTK 838
 Db 803 MQXSTKXESADMDRVERLPKGIYIYVEDMVGK 836

RESULT 5
 AAY81812
 ID AAY81812 standard; protein; 398 AA.

AC AAY81812;

09-JUN-2000 (first entry)

S. pyogenes cysteine protease speB7 protein sequence.

CM Cysteine protease; speB, Group A Streptococcus; extracellular protease;
 KM detection; diagnosis; extracellular matrix; infection; skin infection;
 KM disease status monitoring; vaccine; Streptococcus mediated disease;
 KM pharyngitis; tonsillitis; scarlet fever; sepsis; erysipelas; fasciitis;
 KM pneumonia; acute rheumatic fever; poststreptococcal glomerulonephritis;
 KM cellulitis; bacteraemia; meningitis.

OS Streptococcus pyogenes.

XX US6030835-A.

XX 29-FEB-2000.

XX 16-SEP-1997; 97US-0931220.

XX 02-DEC-1993; 93US-0160965.

PR 14-SEP-1994; 94US-0306542.

XX (BAYLOR COLLEGE MEDICINE.

PI Kapur V, Musser JM;

XX WPI: 2000-205208/18.

DR N-PSDB; AAA07111.

XX Determining the presence of a Group A Streptococcus expressing a
 protease capable of degrading proteins of the extracellular matrix,
 using a specific antibody.

PS Disclosure: Column 7-8; 56pp; English.

CC This sequence represents the S. pyogenes cysteine protease speB7.
 CC The invention relates to a method for determining the presence of a
 CC Group A Streptococcus which expresses an extracellular protease
 CC (preferably speB) capable of degrading proteins of the extracellular
 CC matrix, comprising: (1) combining a sample with an assay medium
 CC comprising a first member of a specific binding pair which binds to a
 CC second member of the binding pair to form a complex, where the first
 CC member has at least 1 epitopic site competitive with at least 1 conserved
 CC epitopic site on the protease; and (2) detecting complex formation as
 CC indicative of the presence of the pathogenic organism. The method is
 CC useful for screening host samples for evidence of infection with
 CC Streptococcus pyogenes and for monitoring the disease status of the host.
 CC The speB gene products are used as a vaccine for protecting against
 CC Streptococcus mediated diseases such as pharyngitis, tonsillitis,
 CC skin infections, scarlet fever, sepsis, erysipelas, fasciitis, pneumonia,
 CC acute rheumatic fever, poststreptococcal glomerulonephritis, cellulitis,
 CC bacteraemia, and meningitis.
 CC Sequence 398 AA;

Query Match 11.1%; Score 494; DB 21; Length 398;
 Best Local Similarity 30.6%; Pred. No. 8,9e-29;
 Matches 129; Conservative 73; Mismatches 155; Indels 64; Gaps 13;

QY 6 LIAIVMLFGIAMOG-----HSAPVTKERALSLARLALROYSLRMGTAVSDKISIDYV 59
 Db 10 LLSLILAGFVLNPNVAFADQNFARNEKEARDSAITFIQKSAIKAGARSAD-1KLDKVN 68
 QY 60 RQDAERGITSQEBGSPATFYVANRGNNEGALVADDRIPITLAYSPIGRFMDSPDN 119
 Db 69 LGSEL-----SGSNMYVYNISTG--GFVYISGDRSPETLIGYSTGSFDAQ--KEN 116
 QY 120 LRMWLOIY-----DOEIGLILSGKAOINEEILRTGCVPAVHAIAMNGHAFANDPMRNO 173
 Db 117 IASFMEYVQIENKRLDITTYAGTAIRQPV-----VKSLLD-----SKGIHQY 162
 QY 174 GYPWNKEPLLPN-----GNHAYTGVATPAQAQIMRYHSMPLQEGSFY----- 218
 Db 163 GNPYNLLTPVIEKVKPGEQSFVCGHAATGCVAATTAQIMRYHNPKNRGLKDYTYTLSSNN 222
 QY 219 -----HAGSLVGNNSGTFGEKMYDMIN--MGNPDDNLNQSDVADATIMRVSASVMSFY 273
 Db 223 PYFNHPKRL--FPAISTROYNMNNTLPYSGRESNVQKM--ATSELMADVGISVDMQY 277
 QY 274 ENGSGTYVYVGLARNNFRYKRSLOLHVRLATYSQEMHDMIRGELASGRPVYAGNNOQ 333
 Db 278 PSSGSGASSRVOAKAKNFQYINOSVHRIKNSDPSKQMEAOIKELSONQPVYQGVGV 337
 QY 334 IGAHAYCDGYASDGTFFHNNMGVSNGFYKLLLSPTSIGIGEGITTYOEITIGIE 393
 Db 338 GGHAFVIDGADGRNFYHVMNGMGVSDGFFRLDALNPSALGTGGAGGFNGYGAVVGIK 397
 QY 394 P 394
 Db 398 P 398

RESULT 6

ABP29579 standard; protein; 398 AA.

AC ABP29579;

02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 8334.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KM group A streptococcus; Streptococcus pyogenes; antibacterial;
 KM antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus pyogenes.

XX WO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB04789.

XX 27-OCT-2000; 2000GB-0026333.

PR 24-NOV-2000; 2000GB-0028727.

PR 07-MAR-2001; 2001GB-0005640.

PA (CHIR-) CHIRON SPA.

PI (GENO-) INST GENOMITC RES.

DR Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

DR Telford H;

DR WPI: 2002-352536/38.
 DR N-PSDB; ABN70210.
 PT New Streptococcus protein for the treatment or prevention of infection

PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
PS Claim 1; Page 3948; 4525pp; English.

CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC Streptococcus (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and anti-inflammatory
CC activity. (I), nucleic acids encoding (II), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
XX Streptococcus proteins.

XX Sequence 398 AA:

Query Match 11.1%; Score 493; DB 23; Length 398;
Best Local Similarity 30.6%; Pred. No. 1.1e-28;
Matches 129; Conservative 72; Mismatches 156; Indels 64; Gaps 13;

QY 6 LLAIVMLFGIIMOG-----HSAPTYKERALSLARLALROVSLRMQOTAVSDKISIDYY 59
DB 10 LLSLALGFGVLANPVFADQNFARNEKEAKDSATFPIOKSAIKAGARSAED-IKLDKYN 68
QY 60 ROGDAERGITSQEGSPAYFYANRGNNGGYALVADDDRLPTLAVSPGFRPDMSPDN 119
DB 69 LGGEL-----SGSNMTVYINISG--GEVIVSDKSPETLGTSGSFDPANG-KEN 116
QY 120 LRMLQIY-----DOEIGLILSGKAOLNEELIRTEGVPAEYHALDNGHFANDPWRWNO 173
DB 117 IASFMSYVEQJKEKKKLDITYAGTAEIKQPV-----VKSLLD-----SKGIHYNO 162
QY 174 GTPWNNKEPLLPN-----GNHAYTCGVATAAQAIRYHSWPLQSGSFDPY----- 218
DB 163 GNPYNLLTFVIEKVRKGEQSFVGOHAAATGCVATATQIKYHNPYKKGDKDYTYLSSNN 222
QY 219 ----HAGSLVGNMSTGFGEMWDIN-MPGPDLDNLTOSQVADYATLMDVSAVSMSFY 273
DB 223 PTFNHNKNL---FAAISTROITNNNTLPTTSGHESVQKK--AISELMADVGISVMDYG 277
QY 274 ENGSGTYVYVYGALRNFRYKRSLOLHVRLATYSQEMHDMIRGELASGRPVYAGNNOS 333
DB 278 PSSGSAGSSRVQALKENFGYNSVHQIRKGFDSKQDWEAOLDKELSQNPVYVQGVKV 337
QY 334 IGHAFVCDGYASDGTFFHFMWGWGCVNGYKLTLLSPSLGIGGEGITTYOETITGIE 393
DB 338 GGHAFVTDGADGRNFYHVMWGWGVSDFRDLALNPASLALGCGAGGNGYOSAVVGIK 397
QY 394 P 394
DB 398 P 398

RESULT 7

AAW07898

ID AAW07898 standard; Protein; 398 AA.

AAW07898;

22-JUL-1997 (first entry)

DE Streptococcus pyogenes clone speB7 pre-pro cysteine protease.

KW Pre-pro; cysteine; protease; inhibition; neoplastic; proliferation;

KW cell; human; treatment; carcinoma; sarcoma; melanoma; lymphoma;
KW leukemia; leukemia; blood; lung; mammary gland; prostate;
KW intestine; stomach; liver; heart; skin; pancreas; brain tissue;
KW wound covering; prevention; metastasis; identification; speB7.

XX Streptococcus pyogenes.

XX Key Location/Qualifiers

FT Misc-difference 216

FT Domain /note="corresponding codon TAG"
333..338
/label= nucleotide_binding_domain

PN M09634941-A1.

PD 07-NOV-1996.

XX 30-APR-1996; 96MO-US05997.

XX 01-MAY-1995; 95US-0432692.

PA (BAYU) BAYLOR COLLEGE MEDICINE.
(TEXA) UNIV TEXAS SYSTEM.

PI Ananthaswamy HN, Fernandez A, Kapur V, Musser JM;

XX MPI: 1996-506148/50.

DR N-PSDB; AAT45219.

PT Use of extracellular Streptococcal cysteine protease enzyme - for
PT inhibiting the proliferation of neoplastic cells, e.g. for treating
PT carcinoma, lymphoma or leukemia.

PS Disclosure: Pages 59-61; 99pp; English.

CC The present sequence is the Streptococcus pyogenes clone speB7
CC pre-pro cysteine protease (CP), which can be used to inhibit
CC neoplastic cell proliferation, especially in a human, useful in the
CC treatment of neoplastic conditions, e.g. carcinomas, sarcomas,
CC melanomas, lymphomas and leukemias originating from blood, lung,
CC mammary gland, prostate, intestine, stomach, liver, heart, skin,
CC pancreas or brain tissue. The CP is especially associated with a
CC wound covering, and can also be used to prevent metastasis or
CC identify susceptible neoplastic cells.
CC K1735 and C519 melanoma cells were injected s.c. into nu/nu mice,
CC optionally followed by i.p. injection of CP (100 microl, 24 hours
CC later). The mice were checked twice weekly for tumour growth for
CC 12 weeks, to give results that showed that treatment with CP
CC completely protected athymic mice against transplanted K1735
CC melanoma growth, and protected 60% of the mice from developing
CC C519 melanomas.

SO Sequence 398 AA:

Query Match 11.0%; Score 489; DB 17; Length 398;
Best Local Similarity 30.4%; Pred. No. 2.1e-28;
Matches 128; Conservative 73; Mismatches 156; Indels 64; Gaps 13;

QY 6 LLAIVMLFGIIMOG-----HSAPTYKERALSLARLALROVSLRMQOTAVSDKISIDYY 59
DB 10 LLSLALGFGVLANPVFADQNFARNEKEAKDSATFPIOKSAIKAGARSAED-IKLDKYN 68
QY 60 ROGDAERGITSQEGSPAYFYANRGNNGGYALVADDDRLPTLAVSPGFRPDMSPDN 119
DB 69 LGGEL-----SGSNMTVYINISG--GEVIVSDKSPETLGTSGSFDPANG-KEN 116
QY 120 LRMLQIY-----DOEIGLILSGKAOLNEELIRTEGVPAEYHALDNGHFANDPWRWNO 173
DB 117 IASFMSYVEQJKEKKKLDITYAGTAEIKQPV-----VKSLLD-----SKGIHYNO 162
QY 174 GTPWNNKEPLLPN-----GNHAYTCGVATAAQAIRYHSWPLQSGSFDPY----- 218
DB 163 GNPYNLLTFVIEKVRKGEQSFVGOHAAATGCVATATQIKYHNPYKKGDKDYTYLSSNN 222

OY	219-----	-TAGSLVGMGSGTFCGEMTDMIN--MPCNPDLNDLTQSQVAYATLMRVSASMSSEY	273
Db	223	PFENIPKNL---FAAISTQIYQWNNILPFTYSGRESNVQK--AISELMDVGIISVMDYD	277
OY	274	ENGSGTYSVYVVALRNFRYKRSIQLHVALYLTQSEWHDMTIGELASGRPVYAGNNOS	333
Db	278	PSSGSAGSSRFVQRLAKENGVNQSHQINRSPFSKDWEMQDKELISQNPVYYQGVK	337
OY	334	IGHAVCCGAYASDGIFFHFMWGMGVSNCGYTKLTLSPTSLGIGSGEIGFTTYGEIITGIE	393
Db	338	GGHAFVVIDGADGRNPFYHVMWGMGVSDGFFRLDALNPALGATGGGAGFNGYQSAVYGTK	397
OY	394	P	
OY	398	P	
Db	398	P	
RESULT 8			
5856			
AAR95856		standard; Protein; 398 AA.	
AAR95856:			
XX	30-OCT-1996	(first entry)	
XX			
XX			
DE		S. pyogenes speB gene encoded extracellular protease.	
XX			
KW		Immunogenic peptide; speB gene; extracellular protease; production;	
KW		antibody; vaccine; diagnosis; detection; Streptococcus infection;	
KW		group A; prevention; treatment; pharyngitis; tonsillitis;	
KW		skin infection; acute rheumatic fever; scarlet fever; probe;	
KW		post-streptococcal glomerulonephritis; sepsis; meningitis;	
KW		erysipelis; cellulitis; fasciitis; toxic shock like syndrome.	
XX			
OS		Streptococcus pyogenes.	
XX			
FE	Key	Location/Qualifiers	
FE	Domain	333..338	
FE		/label= putative nucleotide binding domain	
FE	Region	382..398	
FE		/label= potential collagen docking region	
FE	Misc-difference	216	
FE		/note= "corresponding codon TAG"	
XX			
PN	W03608569-A2.		
XX	21-MAR-1996.		
XX			
XX	13-SEP-1995;	95WO-US11723.	
XX			
XX	14-SEP-1994;	94US-0306542.	
PA	(BAYU) BAYLOR COLLEGE MEDICINE.		
XX			
PI	Kapur V, Musser JA;		
XX			
DR	WPI; 1996-179944/18.		
DR	N-PSDB; AAT15294.		
XX			
PT	Use of extracellular protease(s), partic. cysteine protease - for		
PT	detection, diagnosis prevention and treatment of infection by		
PT	pathogenic organisms, partic. gp. A streptococcus strains		
XX			
PS	Disclosure; Page 12; 97pp; English.		
XX			
CC	The present sequence is the S. pyogenes speB gene encoded		
CC	extracellular protease. An immunogenic peptide derived from the		
CC	protease can be used in the prodn. of antibody (Ab) and vaccine. Ab		
CC	is prepd. by introducing the peptide into a mammal, pref. a mouse,		
CC	followed by Ab isolation. The Ab or a probe derived from the gene		
CC	can be used for the diagnosis and detection of gp. A Streptococcus		
CC	infections, while vaccine, which inhibits streptococcal		

CC	replication, can be used to prevent and treat gp. A Streptococcus
CC	infections, and partic. to ameliorate pharyngitis, tonsillitis,
CC	skin infections, acute rheumatic fever, scarlet fever.
CC	post-streptococcal glomerulonephritis, sepsis, meningitis,
CC	erysipelas, cellulitis, fasciitis and toxic shock like syndrome.
xx	
xx	Sequence 398 AA;
xx	
xx	Query Match 11.0%; Score 489; DB 17; Length 398;
xx	Best Local Similarity 30.4%; Pred. No. 2.1e-28;
xx	Matches 128; Conservative 73; Mismatches 156; Indels 64; Gaps 13;
Oy	6 LLAIVMLGMIAGMOC-----HSAPVTKRALSLRLALROYSLRMGRTAVSDKISIDVY 59
Db	10 LLSLALGGEVLANPVPADQNFARNKEAKDSAITFTOKSAIKAGARSABD-IKLDKVN 68
Oy	60 RQGAERGITSOEGSPAYFYVANKNGNEGVALVAADRIPITLAYSPIGRFMDSPDN 119
Db	69 LGGEL-----SGSNMYVYNISG--GFVIYSGDKRSPELLIGTSGSPDANG-KEN 116
Oy	120 LRMMLQIT-----DQETGLILSGKAQNLNEELRTGVPALVYHALMDNGHFANDPKRMNQ 173
Db	117 IASFMESEYVEQIKENKKIDITTYAGFAEIKQPV-----VKSILD-----SKGIHYNQ 162
Oy	174 GYPNNKKRPELLPN-----GNNAVTCVATAAOLIMRYHSMPLQEGSPDY----- 218
Db	163 GNPNYLLTPVIEKKYKPGQSFVGQALAAQCVATAAQIMKKYHNYPNKKIKDITTYLSSNN 222
Oy	219 ----HAGSLVNMWSGTFCGEMTDWIN-MPGNDPLDLNLTQSOVDAYATLNRDVSASVMSFY 273
Db	223 PYFNHPKRL-----FAISTRQINNMNNILPTTYSGRESNVQKM--AISELMADNGISVDMQYG 277
Oy	274 ENGSGTYSVYVVALRNFRYKRSLOLHVRLALYTSQEMHDMIRGELASGRPYVYAGNNQS 333
Db	278 PSSSAGASSRVQALKEKFRGYNQSVHQINRSDFSKODWEAQIDKELSONQPPYVYOGVGKV 337
Oy	334 IGHAFVDCGYASDGFHFHNMWCGVSNFFYKITLLSPFSLGIGEGSIGFTTYQELITGIE 393
Db	338 GGHAFVLDIGADGRNPFYHVMWGMGVSDGFRLDALNPALGTGGAGGPFNGYSAAVGTK 397
Oy	394 P 394
Db	398 P 398
xx	
xx	RESULT 9
xx	AAB36098
xx	ID AAB36098 standard; Protein; 398 AA.
xx	AAB36098;
xx	16-FEB-2001 (first entry)
xx	Streptococcus pyogenes streptopain.
xx	Streptococcus pyogenes; streptopain; periodontain; antiinflammatory;
xx	antibacterial; amidolytic; alpha_1-proteinase inhibitor; periodontitis;
xx	gingivitis.
xx	Streptococcus pyogenes.
xx	W0200063394-A2.
xx	26-OCT-2000.
xx	20-APR-2000; 2000MO-US10574.
xx	21-APR-1999; 99US-0130436.
xx	(UYGE-) UNIV GEORGIA RES FOUND INC.
xx	(TRAV/) TRAVIS J.
xx	(POTE/) POTEMPA J.
xx	(NELS/) NELSON D.

DB 400 NKDLMSDFTYTSANALYTPQVD-----INPTKGLVCSWVEYLPGRKI 444
QY 753 Y--TYKIVNGTAVEALESSE-EIRVFPNARDYVEISAPCIPOETSIIILFDLSGKIYMN 809
DB 445 VMSDTQMTANGVEDIVMOEGSMKLYPNPAQEAIVISLP-TAANCAKAVYDMDGRIYVAA 503
QY 810 SLGAGHRDVSRLPNCAYILKY 832
DB 504 SFGSGNEYRLNVOHLAKGTIILKY 526

RESULT 11

AA34430

ID AA34430 standard; Protein: 563 AA.

AC AA34430;

25-AUG-1999 (first entry)

Porphyromonas gingivalis protein PG91.

Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
vaccine; antigenic.

Porphyromonas gingivalis.

W09929870-A1.

17-JUN-1999.

10-DEC-1998; 98W0-AU01023.

04-AUG-1998; 98AU-0005028.

10-DEC-1997; 97AU-0000839.

31-DEC-1997; 97AU-0001182.

30-JAN-1998; 98AU-0001546.

10-MAR-1998; 98AU-0002264.

09-APR-1998; 98AU-0002911.

23-APR-1998; 98AU-0003128.

05-MAY-1998; 98AU-0003338.

22-MAY-1998; 98AU-0003654.

29-JUL-1998; 98AU-0004917.

(CSLC-) CSL LTD.

Agus CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
Ross BC, Rothel LJ, Webb EA;

WPI: 1999-385613/32.

N-PSDB: AA391648.

Antigenic Porphyromonas gingivalis peptides for preventing
gingivitis

Claim 1; Page 404-405; 588pp; English.

AA391536 to AA391801 encode two hundred and sixty six antigenic
Porphyromonas gingivalis (Pg) polypeptide sequences given in AA394318 to
AA394533. AA391802 to AA391989 represent PCR primers used in the
isolation of the Pg polypeptides. The Pg polypeptides have antibacterial
activity with a vaccine mechanism of action. The Pg polypeptides can be
used as vaccines especially against Porphyromonas gingivalis. Probes can
be used to detect Porphyromonas gingivalis in standard hybridisation
assays. Porphyromonas gingivalis is involved in periodontal disease
especially gingivitis.

Sequence 563 AA;

Query Match 4.3%; Score 190; DB 20; Length 563;

Best Local Similarity 20.5%; Pred. No. 1.4e-05;

Matches 103; Conservative 78; Mismatches 162; Indels 160; Gaps 23;

QY 413 DIEAEKSESGLVN-----GX-----SIYNTGEQSNLDLGRKNADEVEIEVT 458
DB 124 DIVYTKNEDSDIKIMVELMKNRPGYKSKVAAPSRANNAONAKLYKEPNSQILDVVI 183
QY 459 SSINISWYIGGEHPESESLAPNOLSGINTITLLYRTGTEQWEPVRAHAGGVNSIKVN 518
DB 184 AS-----NY-----RSPSLNNGGNPFALAFAYTGF-----209
QY 519 TDDPNVYVYVONNECKLSTVPSFVADNYSYHSITTYQFNSDPDELRTVAFALSTG 578
DB 210 -----NTHKISFV--DYFSLNGGONFNKLLFSQDGKKID--KVDLSLG 252
QY 579 ATADYISLGWYAEV-----PGGSSNTPVY-----WS-----K 607
DB 253 STSESGHNAAPLMGVFEMNKGKSDIGFLSNFVNDPEPQMSGPIKVSSEDSMFSR 312
QY 608 DVLTLESGDYTL---WYRFSI-----NNQDEMKKIGSVSKTPPEYTHPLEVGHNOT 658
DB 313 IOMLDEDDNNTINGESCHNFMITYSDYDEYSDW-----DIRVYYPKSKFYK 362
QY 659 STYTLDMANNRVLPDF--TLKNGLPF-----NGLVYVFRQOSSSGSL--WAAQETVH 709
DB 363 KPTMDLVEALFTASQSETNSGLGYDKNANHLYLTIVAKKEENGNTLYKRWANVDKH 422
QY 710 IKO--GETP-----VYKPYVEGPIPDGSYRATLHAFFVNGQOL-----YLGKRN 752
DB 423 NKDLMSDFTYTSANALYTPQVD-----INPTKGLVCSWVEYLPGRKI 467
QY 753 Y--TYKIVNGTAVEALESSE-EIRVFPNARDYVEISAPCIPOETSIIILFDLSGKIYMN 809
DB 468 VMSDTQMTANGVEDIVMOEGSMKLYPNPAQEAIVISLP-TAANCAKAVYDMDGRIYVAA 526
QY 810 SLGAGHRDVSRLPNCAYILKY 832
DB 527 SFGSGNEYRLNVOHLAKGTIILKY 549

RESULT 12

AA34522

ID AA34522 standard; Protein: 921 AA.

AC AA34522;

25-AUG-1999 (first entry)

Porphyromonas gingivalis protein PG57.

Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
vaccine; antigenic.

Porphyromonas gingivalis.

W09929870-A1.

17-JUN-1999.

10-DEC-1998; 98W0-AU01023.

04-AUG-1998; 98AU-0005028.

10-DEC-1997; 97AU-0000839.

31-DEC-1997; 97AU-0001182.

30-JAN-1998; 98AU-0001546.

10-MAR-1998; 98AU-0002264.

09-APR-1998; 98AU-0002911.

23-APR-1998; 98AU-0003128.

05-MAY-1998; 98AU-0003338.

22-MAY-1998; 98AU-0003654.

29-JUL-1998; 98AU-0004917.

(CSLC-) CSL LTD.

Agus CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
Ross BC, Rothel LJ, Webb EA;

QY 605 WSKDVLTLSEGDYTLWRESINNOXDEMKKIGS-----VSKTPTEY-THPLFEVG--- 654
 Db 605 -----DYPPTMTG-----HSEKCIASPSYLPIMIGVLPNTLVTPRLLEGAKLV 649
 QY 655 ----HNQTSYTYLDMAHNRVLPDFTLKNLGLPENGELVYVFPOTOSSSGSLMAAQTVAH 710
 Db 650 KYWVSAQDAVYSAE--HYAVMVST-----GTAVE-DFVLLFEETMTAKANGMYERTITL 702
 QY 711 KQGETF-----VYKPVVEGPIPDGSYRATL----- 735
 Db 703 PACTKIAMRHVCTDMFELLDDITYVSTETVPEVYDFVSLIENKGRLLKNYPNG 762
 QY 736 -----HAFVNGQOOLYLK----- 748
 Db 763 YEPDKTDDKPLQLTGYNITVANGSLVHIQDPVLEYIDETYSRRDQVEMECYTAAYN 822
 QY 749 -----GKRNVTYKIVNGTAVEAIESESEIRVFPNPARDVVEISAPCIPOETSIILF 799
 823 DNIESOSVCDKLNVTI-----TSLDNIQSDTSIKITYPNPASVYVREG-LSRSKSTIELY 876
 800 DLSGKIYMNKNSLSAGHGRMDVSRPLPNGAVILKYDG---YTTKINI 841
 877 NALGICILREETHSEKTEIDVSRNDGQVYLKIVGKNKTTKEVEI 922
 Db

RESULT 15

AAV34392
 ID AAV34392 standard; Protein: 938 AA.

XX AAV34392:

DT 25-AUG-1999 (first entry)

XX Porphyromonas gingivalis protein PG57.

XX Porphyromonas gingivalis; PG: periodontal disease; gingivitis;

XX vaccine; antigenic.

XX Porphyromonas gingivalis.

PN W09929870-A1.

XX 17-JUN-1999.

XX 10-DEC-1998: 98MO-AU01023.

XX 04-AUG-1998: 98AU-0005028.

XX 10-DEC-1997: 97AU-0000839.

XX 31-DEC-1997: 97AU-0001182.

XX 30-JAN-1998: 98AU-0001546.

XX 10-MAR-1998: 98AU-0002264.

XX 09-APR-1998: 98AU-0002911.

XX 23-APR-1998: 98AU-0003128.

XX 05-MAY-1998: 98AU-0003338.

XX 22-MAY-1998: 98AU-0003654.

XX 29-JUL-1998: 98AU-0004917.

XX (CSLC-) CSL LTD.

XX Agius CT, Barr IG, Hocking DM, Margetta MB, Patterson MA;

XX Rosa BC, Rochel LJ, Webb EA;

XX WPI: 1999-385613/32.

XX DR N-PSDB: AAX91610.

XX Antigenic Porphyromonas gingivalis peptides for preventing

XX gingivitis

XX Claim 1: Page 361-363; 588pp: English.

XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic

XX Porphyromonas gingivalis (PG) polypeptide sequences given in AAV34318 to

CC AAY34583. AAX91802 to AAX91889 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.
 XX
 SQ Sequence 938 AA:

Query Match 3.6%; Score 158.5; DB 20; Length 938;

Best Local Similarity 17.7%; Pred. No. 0.0075;

Matches 146; Conservative 111; Mismatches 242; Indels 327; Gaps 37;

QY 221 GSLVNMSTGFGBMYDWINP---GNPDIDNTQSQVDAVATLMDVSAVSMEFENG 277
 Db 232 GQNGVGRLLNNYEDY-----QPEGKGNEL-QLSGVITVANGLLAQK---DVSILEYVD 283
 QY 278 GTYSYVVGALRRNFRKSLDLH--VRALYTSQEMDMIRGLASGRVYAGNQSIG 335
 Db 284 STYS-----LRDN-----PLQVEYCVTAIV-----DESIE 308
 QY 336 HAFVCD--GYASDGFHFNMGGVNGFYKLLTSPSTLIGIGEGIFTYIEITIGIE 393
 Db 309 SSTVCGTLHAYDAILYENFENGVPVNGM-----LVLDAGDGRSMGHYL----- 353
 QY 394 PAKTPAAGTDLPLIALKDIEAEYKSEGLNLYG---SIYNTGEEQSNLDLGYRLNKA 449
 Db 354 -----NAYDAFP-----GHNGHCSLSASVYVPGIGVTPNPVITTPRV 391
 QY 450 DG-EVLEKTSINISWYGEHPESFSLAPNOLSGITITLLKRTTEQ---W--E 502
 Db 392 EGAKRVKVVWSTODAMAA---EHYAVMASTTGAVDFFVLEETMTAKPTGAWYER 446
 QY 503 PVRAAG--GYVNSIKVNTD-----PNNVYVVDNNEKLS-I 538
 Db 447 TNLPEGTIYAMRHVNCITDIYFLKADITVGTAPSEPEVYDFVSLIENKGRLLKN 506
 QY 539 VPNSF-----VADLNSYEHSTITVOFNSDSPDIRTPVAFALSTGATADVIS 586
 Db 507 YPNGYEPDKTDKDPQLAGYNITVANGSLVHIQDP-----TVLEYIDETYSRRDQVE 560
 QY 587 L-----GVMAEVPGGSSNYPVV 604
 Db 561 VEYCVTAAYVNDNIESOSVCDKLIYDSQSDIILYEGFEAGSIPEGWLLDADGDVNMV--- 617
 QY 605 WSKDVLTLSEGDYTLWRESINNOXDEMKKIGS-----VSKTPTEY-THPLFEVG--- 654
 Db 618 -----DYPPTMTG-----HSEKCIASPSYLPIMIGVLPNTLVTPRLLEGAKLV 662
 QY 655 ----HNQTSYTYLDMAHNRVLPDFTLKNLGLPENGELVYVFPOTOSSSGSLMAAQTVAH 710
 Db 663 KYWVSAQDAVYSAE--HYAVMVST-----GTAVE-DFVLLFEETMTAKANGMYERTITL 715
 QY 711 KQGETF-----VYKPVVEGPIPDGSYRATL----- 735
 Db 716 PACTKIAMRHVCTDMFELLDDITYVSTETVPEVYDFVSLIENKGRLLKNYPNG 775
 QY 736 -----HAFVNGQOOLYLK----- 748
 Db 776 YEPDKTDDKPLQLTGYNITVANGSLVHIQDPVLEYIDETYSRRDQVEMECYTAAYN 835
 QY 749 -----GKRNVTYKIVNGTAVEAIESESEIRVFPNPARDVVEISAPCIPOETSIILF 799
 Db 836 DNIESOSVCDKLNVTI-----TSLDNIQSDTSIKITYPNPASVYVREG-LSRSKSTIELY 889
 QY 800 DLSGKIYMNKNSLSAGHGRMDVSRPLPNGAVILKYDG---YTTKINI 841
 Db 890 NALGICILREETHSEKTEIDVSRNDGQVYLKIVGKNKTTKEVEI 935

RESULT 16

AAG98256

PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 XX biodiversity

Claim 20; SEQ ID No 42106; 103bp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp://ftp.int/pub/published/pt_sequences.
 XX

Sequence 2652 AA;

Query Match 3.48; Score 153; DB 22; Length 2652;
 Best Local Similarity 17.88; Pred. NO. 0.1;
 Matches 186; Conservative 145; Mismatches 334; Indels 382; Gaps 50;

46 QTAASDKISIDYVROG-DAERGITSQEGSPAFYVANGN-----NEG 90
 1684 KTAKGKGYCTVLTLPDGTAEADVINEGOTIDIFTAKPGTYIVYVFGGVDPNSPF 1743
 91 ALVAADRIPTILATPAGIFGDMSPDNLRLML-----QYDEIGLI 134
 1744 TVMATGEV-TAVEAPV-----MACPGFPMVTEAVPVSDMNGIGCFKFDLVPEA 1797
 135 L-----SGKAQNEILRTG-----VPAEV-----HALDNGHFANDPMRW 171
 1798 VRKGEITGEVHMPGSKTATPEIVDKDGTIVRYAPTEVGLHEMKIKMGSHIESPLGF 1857
 172 NOGYPMNNK-----EPILPNG--NNAVTCGVAATAAQIMRYHSMPLGEGSFIDYHAG-- 221
 1858 YVNYPMSSGSVAYGGLVYGVANKTAFTTIEDA-----GEGGDLAIEGFSK 1906
 222 ---SLVGMMSGTF-----GE-----MYDWINMFGNP-----DLNLTQSOVDAY 257
 1907 AEISCIDKDKDCTVYVLTPLPGDYSLIVKYNOKHIGSPFTAKITDDSRCSQVRLGSA 1966
 258 ATLNRDVS-----ASVSMSPFENGSGTYSVYVGAALRNFRYKRSLOLHYRALTTSEMD 313
 1867 ADFLLDISEIDSLTAST-KABSGRDEPCLRLPNN-----HIGISFIRENGE 2016
 314 -----MIRGELASGRPVYVYAGNNSIGHAFPCDGTASDGFHEMW 353
 2017 HLVSIRKKNHNVANPVSIMVQSEIGDARRARYVGRGLSEGRTEMSDFIVDR--DA 2073
 354 GMGCVS-----NGFYKLTL-----SPISLGIG 376
 2074 GYGCISLAEVGSKDIDEDLEDCCKCYEPFPGVYIVSTRKADHVGSGSPFYKIS 2133
 377 GEGIGTITIOETITGIEPAKTPAEAGTDA-----LPIALDIEAEKESGGLNVGYSIY 431
 2134 GEG-----RVKESITRTSRAPVATVGSICDLAKIPETINSSDMSAHYTSBSG----- 2181
 432 NTGEGQSNLDGRLANKADGEVIEKYSINISWYGEHPESESPLAPNLDGGINITL 491
 2182 -----RVTEA--EIVPM-----GKNSHCVRFPDEK--GVHTVSV 2212

492 LVR-----RTGTEQWEP-----VRHAGGYVNSI 515
 2213 KYRGQHTGSPFOPTVPLDEGGAHKVRAGGPELLEGEGVPAEISWIREAGAG---GL 2269
 516 KVNITDDNNVYVTVDNNE-GKLSIVPNSFVA-DLNSYEHSITVOPNSDSDPEIRTPVAF 573
 2270 SIAGEVSKAEIIFDDHKNGSCGV---SYIAGEPQWTE--VSIFEN--DEHIESPY 2319
 574 ALSTGATADVVISLGMVMAEVRGSSNYVYVMSKVDLALSEGDTLMTYFRSINNOKDEWK 633
 2320 LVPVIAPSDARLITWMSIQESGLKVNOPAS-----FAILNKA----- 2359
 634 KISSVVKPTETVTEHLEFVGNHNTSTYTLDMANRVL-PETLKNGLPENGELV--- 689
 2360 --GKIDAKVHS-----PSGAVECHVSELEPPKXAVRFLPHENGVTIDVFNESHVVGSP 2413
 690 ---VFROTQSSGSLMAAQETVHIKQETVYRP-----VREGP----- 725
 2414 FVVRVGEQGAQNPALVSAVGT-GLEGGTGIOSEFFINTTRAGPGLTSTYIEGSPKVM 2472
 726 ---IPDSYRATLHAFVNGOOLYLK-GKRNVTYKIVNGTAVAELESSEIRVFPNPAR 780
 2473 DCQETPEG-YKVITYTPMARGNYLISVAYGPNH---IVGKSPKAVTGQRL----- 2520
 781 DYVEISAPCIPOETSIILFDLSGKIYVK-----NSLSAGHR----- 817
 2521 ---VSPGSAPNETSSILVESVTKVYKRYVXAROPKGIQPNKVTSGXGGSQKAEV 2576
 818 ---MDVSRLEPNGAYILKVDGYTT 837
 2577 GQKSSFLVDCSKAGSNMLLIGVHGPTT 2603

RESULT 18
 AAY34479
 ID AAY34479 standard; Protein; 821 AA.
 XX
 AC AAY34479;
 XX
 DT 25-AUG-1999 (first entry)
 XX
 DE Porphyrymonas gingivalis
 XX
 KW Porphyrymonas gingivalis; PG; periodontal disease; gingivitis;
 XX vaccine; antigenic.
 OS Porphyrymonas gingivalis.
 XX
 PN WO9929870-A1.
 XX
 PD 17-JUN-1999.
 XX
 PF 10-DEC-1998; 98MO-AUD1023.
 XX
 PR 04-AUG-1998; 98AU-0005028.
 PR 10-DEC-1997; 97AU-0000839.
 PR 31-DEC-1997; 97AU-0001182.
 PR 30-JAN-1998; 98AU-0001546.
 PR 10-MAR-1998; 98AU-0002264.
 PR 09-APR-1998; 98AU-0002911.
 PR 23-APR-1998; 98AU-0003128.
 PR 05-MAY-1998; 98AU-0003338.
 PR 22-MAY-1998; 98AU-0003654.
 PR 29-JUL-1998; 98AU-0004917.
 XX
 PA (CSLC-) CSL LTD.
 XX
 PI Agius CT, Barr IG, Hocking DM, Margetis MB, Patterson MA;
 PI Ross BC, Rothel LJ, Webb EA;
 XX
 DR WPI: 1999-385613/32.
 DR N-PSDB; AAX91697.

XX Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis
XX
XX
PS Claim 1; Page 460-461; 588pp; English.

CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (Pg) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91899 represent PCR primers used in the
CC isolation of the Pg polypeptides. The Pg polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The Pg polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gingivitis.

Sequence 821 AA:

Query Match 3.3%; Score 148.5; DB 20; Length 821;
Best Local Similarity 19.3%; Pred. No. 0.035;
Matches 151; Conservative 87; Mismatches 249; Indels 297; Gaps 37;

176 PMNKEPLIPNGNHAYTCVATAAQMRYHS-----W----- 208
176 PMNKEPLIPNGNHAYTCVATAAQMRYHS-----W----- 208
216 PLTNPDGAYRAGNHVVOGAT-----RYNANNVDLNRNFKDDVAGDHPDGKRWQPEAT 267
209 ---PLOGEGSF---DYHAGSLVGNMSTGTEGEMDWINMGPNDLNLQSOVD----- 255
268 AFMDLEGNTSFVLGAINHGTEYVN-----YPM-----DNKKERRHADDEWYKL 310
256 ---AVATLMRDVSAVSMSFYENGSGTYS---YVVVGALR---NNFRYKRSIQLHVRAL 305
311 ISRNVAACQASISASVMTS--ETNSGILNGSDMYVIRGSDQANAFHRLRETTLEI--- 365
306 YTSQEHMDIRGELASGRPVYTAGNNSGIGHAFVCCGYASDGTFFHNMWGVSGNGFYKL 365
366 ---SNMKLVPAQLPRKYMNLNKESTL-----L 388
366 TLLPSTSLGIG---EGIFTYQETITG---IEPAT----- 397
389 ALIEESLYGHIHGVTSANGOPKLCQILLENHDKRSDVSDTGTGYAPITAGYTVK 448
398 -PAEACTDALPILALDIDAEVYKSES-GLNVGSIYNTGEGSNLDLGYRLNKADGEVIE 455
449 YKAEGYPEATRTTIKDEKIVIMDLGNSVPLPVPDPTASPTISGEGVQPOD---Q 504
456 VRTSSINISWYCGEPESESLAPNOLSGCINTITLIRRTGTEQNEPVRHAAO-GGYVNS 514
505 TTNMPTNMEMTEFGGQP-----AMSTEONPLVSYSHRGQYDVT 542
515 IKV-NTTDPNNVY---VTVDNNEKSLIYV-NSFVADLNSYHSHITTVQFNSDPEIR 568
543 LKVMANSGSNTTIKEKFTIVN-----AVMPVAEFGTPEIEGO-TVSFQNSGTAIN 595
569 TPVAFALSTGATADYISLGWVAEYVGGSSNTPVWYSKDYDLTSLSGDYTLRFSSINQ 628
596 YWIFEGGTPATSESD-----ENPTVLXSK---AGQYDVL----- 627
629 KDEMKAIGS--VSVKPTLEYTHLFEVGNHQSITYLDMANHNVLPDTIKNLGLPENG 666
628 ---KAISASGETVKTREY-----ITVKKAPVP---A 653
687 LVVVFQOTSSSSSLAAOETVHIKGEFFVYK-----VEGRIIP----- 727
654 PVADFGTIRK-----VKKGEVTFERDLSTNNFTSLWLFEGGSPATSTEQNP 701
728 ---DGSYRATLHAFAVNGOOLYLKGRNVTYKIVNGTAVAISSSEIRFPPNPAR 781
702 VVTYNTEKRDVOLTNTNBSGN--VKKADY-IEVILDSVEDIYAQGIIVIRPONGTK 758
782 YVEISAPCIPOETSIILFDLSKTIWKNLSLASHGRADVRLPENGAVILKV---DGYTT 837
759 QILIEANNAIK--AIVLYDINGRVVLKTPNQLRSTVDLSLDEGATITINIKREKSARTE 816

QY 838 KINI 841
DB 817 KINI 820

RESULT 19
ID AAY34354 standard; Protein; 869 AA.
XX
XX AAY34354;
AC
XX 25-AUG-1999 (first entry)
DT
XX
DE Porphyromonas gingivalis protein PG21.
XX
XX Porphyromonas gingivalis; Pg; periodontal disease; gingivitis;
KW vaccine; antigenic.
XX
OS Porphyromonas gingivalis.
XX
XX WO9929870-A1.
XX
XX 17-JUN-1999.
XX
XX 10-DEC-1998; 98WO-AU01023.
XX
XX 04-AUG-1998; 98AU-0005028.
XX
XX 10-DEC-1997; 97AU-0000839.
XX
XX 31-DEC-1997; 97AU-0001182.
XX
XX 30-JAN-1998; 98AU-0001546.
XX
XX 10-MAR-1998; 98AU-0002264.
XX
XX 09-APR-1998; 98AU-0002911.
XX
XX 23-APR-1998; 98AU-0003128.
XX
XX 05-MAY-1998; 98AU-0003358.
XX
XX 22-MAY-1998; 98AU-0003654.
XX
XX 29-JUL-1998; 98AU-0004917.

XX (CSLC-) CSL LTD.

XX Agius CT, Barr IG, Hocking DM, Margelits MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;

XX -WPI; 1999-385613/32.
XX N-PSDB; AAX91572.

XX Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis
XX

PS Claim 1; Page 316-318; 588pp; English.

CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (Pg) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91899 represent PCR primers used in the
CC isolation of the Pg polypeptides. The Pg polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The Pg polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gingivitis.

Sequence 869 AA:

Query Match 3.3%; Score 148.5; DB 20; Length 869;
Best Local Similarity 19.3%; Pred. No. 0.038;
Matches 151; Conservative 87; Mismatches 249; Indels 297; Gaps 37;

176 PMNKEPLIPNGNHAYTCVATAAQMRYHS-----W----- 208
264 PLTNPDGAYRAGNHVVOGAT-----RYNANNVDLNRNFKDDVAGDHPDGKRWQPEAT 315
209 ---PLOGEGSF---DYHAGSLVGNMSTGTEGEMDWINMGPNDLNLQSOVD----- 255

Db 316 AFMDLEGNTSFVLGNHIGTEYVN-----YPM-----DNKERRHADDEMYKL 358
 Qy 256 ---AVATLMRDVSVSISFEYENGSGTYS---VYVVGALR-----NNFRYKRSQLOLHVRL 305
 Db 359 ISRMYAAACOSISASYSMTS---ETNSGIINGSDMYVINGSRQDNANFHLRETTLEI--- 413
 Qy 306 YTSOEHMDMIRGELASGRPVYAGNNOISGHAFVCGVYSDGTFHNMKGVSNGCYKL 365
 Db 414 -----SNTKLVPASQOLPKYVNLNKESL-----L 436
 Qy 366 TILSPSTSLGIG-----EGIGFTIYOELING---IEPAKT----- 397
 Db 437 ALIEBSLYGHVITVSAANCOPKCOILLIENHDKRNSDVYSDATGYVYRPKAKATYVYK 496
 Qy 398 -PAEAGTDLALILALDIEAEYKSES-GINVGSIYNGEBSNDLGRKLKADGEVIE 455
 Db 497 YKAGYPEARITITIKDEYVIMDIALGNSVPLPPDPTASMTISVESVOFOD---Q 552
 Qy 456 VKTSSINISWYGEHGEHESFSLAPNOLSGINTITLLYRRTGTEQMEFVRHAQ--GGYVNS 514
 Db 553 TTNPNPTNMEFEGGQP-----AMSTEQNPLVSYSHPGQYDYT 590
 Qy 515 IYV-NTTDPNNVY---VTVDNNEKLSIVP-NSFVADLNGYEHSTITVOFNSDPDEIR 568
 Db 591 LKVMNASGSNTITKKEFTIVN-----AVMPVAEPVCGPTIEBGO--TVSFQONSTVATN 643
 Qy 569 TPVAFALSTGATADVISLGVMVAFVPGSSNTPVNSKDYLTLSBGDYTLMYRFSINQ 628
 Db 644 YVIMPDGTPRASED-----ENPTVLVSK---AGQYDVTL----- 675
 Qy 629 KDEMKKIS--VSVKPTETHTPLFEVGHNOTSTYTLDMANRRLDPFLKNGLPRNGE 686
 Db 676 ---KAISASEETVYKTEKY-----ITVKKAPVP---A 701
 Qy 667 LVVVFROTOSSSGLMAQETVHIKOGETVYK-----VVEGPVP----- 727
 Db 702 PVADFEGRPRK-----VKGEYTFKDLSTNNFTSWLWMEFEGGSPATSTEQNP 749
 Qy 728 -----DGSYRATLHAFVNGQOOLYLKGRNRYTVKIVNGAVAEIASSEIEIVFRPARD 781
 Db 750 VVTYNETGKYDVLATNEGGSN--VKKADY-IEVILDDSVEDIYAQGTGIYIRPONGTK 806
 Qy 782 YVEISAPCIQETSIILFDLSGKIYKNSLSAGHGMVSRLLPNCAYILKY---DGYTT 837
 Db 807 OILIEANAAIK--AIVLYDINGRNVLTTPNOLRSTVDLSLPEGIYITINIKTEKSARTE 864
 Qy 838 KINI 841
 Db 865 KINI 868
 RESULT 20
 AAG63962
 ID AAG63962 standard; Protein; 1352 AA.
 AC AAG63962;
 XX 29-OCT-2001 (first entry)
 DE Amino acid sequence of xyloglucanase enzyme.
 XX xyloglucanase; family 44; glycosyl hydrolase.
 KM cellulosic fiber; textile scouring.
 OS Paenibacillus polymyxa.
 PM MO200162903-A1.
 PD 30-AUG-2001.
 PF 21-FEB-2001; 2001MO-DK00116.
 PR 24-FEB-2000; 2000DK-0000291.

XX (NOVO) NOVOZYMES AS.
 PA Schmitt K, Jorgensen PL, Schuelein M;
 PI WPI: 2001-552819/57.
 DR N-PSDB: AAH75059.
 XX
 XX New xyloglucanase enzyme belonging to glycosyl hydrolases family,
 PT useful for detergent compositions, and textile or cellulose fiber
 PS processing industries
 PS Example 3; Page 81-85; 97pp; English.

CC The present sequence represents a xyloglucanase of the invention. The
 CC specification describes a xyloglucanase enzyme belonging to family 44
 CC of glycosyl hydrolases and exhibiting a relative xyloglucanase activity
 CC of at least 30% at pH 5-8. The enzyme exhibits high performance in
 CC detergent compositions and prevents binding of certain soils to the
 CC xyloglucan left in the cellulosic material. It is stable at pH 5-10 at
 CC room temperature and has a half life of more than 50 days when incubated
 CC in a full formulated liquid detergent at 30 plus degrees celcius. The
 CC enzyme is used in detergent compositions, textile industry for improving
 CC the properties of cellulosic fibers, yarn, and woven or non-woven
 CC fabrics, preferably in textile scouring process, and in cellulose fiber
 CC processing industry for rattling of fibers e.g. hemp, jute, flax, and
 CC linen.

Sequence 1352 AA;

Query Match 3.38; Score 148.5; DB 22; Length 1352;
 Best Local Similarity 18.68; Pred. No. 0.076;
 Matches 176; Conservative 127; Mismatches 334; Indels 307; Gaps 47;

Qy 112 DMSMPNDLRLMQLIYDOEIGL-----ILSGKAQLEILLRTGCV 151
 Db 282 DMSDVKNYSMPFYDYLIDKRLSSQYEGKRLDVEDVHYPRAMGGIRITNEVGNDETK 341
 Qy 152 PAVYHALMDNGHRAANDPKNNGY-----PMNNK-EPILP----- 185
 Db 342 KAKQA-----PRLMDPYKEDSWIAQWNSPELILPRLQSVDKYPPGKTLAMT 392
 Qy 186 ---NGNHAYTGCVATAAA-----QIMRYHSPLO--GESEFDYHAGSLVGNMG--T 230
 Db 393 EYSYGEENDISGSIANTDVLGLKNDVYMAVYKRLKGVNNYSAAYKTLRNDGKNS 452
 Qy 231 FG-----EMYMINKPGNPDDNLTOSQVAYATIMRDVASYSMSFYENGSGTY-SYV 283
 Db 453 FGDTSVSAQTSDIVNSVHASVYTNASDKELH-LVWNKMSMDASADADPDLSGANTYISGK 511
 Qy 284 VVGALRNNEFRYKRSQLOLVY---RALYTSQEHMDMIRGELASGRPVYAGNNOIS---GH 336
 Db 512 VMGFDKNSQIKREAPITIOISGNRTYVP-----PLTAYHIVLTGNDTSPEGPE 563
 Qy 337 AFVCDGYASDGTFFHNM-----GMG-----GVNKG-- 361
 Db 564 SEKLAEGDGKAVHLSWDASSGVYGVQVATDENGPFVAVSNLLETSTTDTNVTNGTS 623
 Qy 362 -FYKLTLSPTSLGIGEGIGFTIYOELINGIPAKTPAAGADALPILALD--TEAY 418
 Db 624 YYYKYV--AKTNKSSSSNLLKAVPKMPVNG--PAREVEEGT-----LKGTYESSG 672
 Qy 419 KSEGLWNGYSIYNTGEQS-----NLDLGR-----LN-KAAGEVIE 455
 Db 673 TGISGAGYVTFNHNPNQSLTMTIQAPFAGLYLITIGRPHDKNRNSLNAKAGAGELLL 732
 Qy 456 VKTSSINISWYGEHGEHESFSLAPNOLSGINTITLLYRRTGTEQMEFVRHAQGYVNSI 515
 Db 733 KKTADF-----KETSQKYLNLNAGANTISF--ETGM-GMYDI-----DYVRL 772
 Qy 516 KVNTTDPNNVYVTVDNNEKLSIVPNSFYAD-----LNSYEHSTITVOFNSDPDEIR 568
 Db 773 PAADRPBHAVYTKLTN-----PNAIVEAKALNNIVLDQYGNKMSLGSQEIINEIDMLO 824

```

QY 569 TPVAFALSTGATADVITSLGWMAE--VPGSSNPVWNSKD--VLTISEGDYTLWY--- 621
DB 825 ANV--GKKPAIALDLIDYSPSRAEHLSTEAERAIANDKOGIVTFA-----WHMNA 876
QY 622 -RFSINNO-KDEWKKIGSVSVKPTPEYTHPLFEVGHNOTSTYTLDMANHRVLPDFTLKNL 679
DB 877 PKGLIDTQCKEMWRGRFYADSTFEDIEYA-----MNHPESEDYKLLIRIDIDYIAG-QLKKL 930
QY 680 GLPFNGELVYVFRQTOSSSGS--LMAAQETVHIKQ----- 712
DB 931 Q--DAKYVPLFRPLHEAGKMFWMGAKGPEPYKKLYIIMHDRLTNVHKLNLIWVWNSY 987
QY 713 -----GETFV-----YKPVY-----EGPIPD-----G 729
DB 988 APDWTFPGDEYVDILSFDSTFPAQADYSPQISKYEDLVALGDKKLVAMSENGPLPDPDLMK 1047
QY 730 SYRATLHAFFVNGOOOLYLKGRH--YTVKIYNGTAVEAIESSEIIRVFPNPARDYVEIS 786
DB 1048 AYQAHWSMFATWYGDFVRDCKQNSLEHLKKYVNHPRVITLDE-----LPTNKTGITE 1101
787 APCIPQETSIIIFDLGKIYMKNSLSAGHRMDYSR--LPNGAY 828
1102 QPSVPGSFTLNAAGETAKVLSLSTASANKSYEVKRSTENGAF 1145

```

Search completed: May 29, 2003, 13:25:39
 Job time : 48 secs



•
•
•

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 13:24:52 ; Search time 18 Seconds
(without alignments)
1377.974 Million cell updates/sec

Title: US-10-030-330-1

Perfect score: 4438
Sequence: 1 MKKSFLLATVLMFGIMQGH.....PNCATILKVDGYTTKINIYH 843

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 segs, 29422922 residues

Number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents A.A.
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/prodata/1/1aa/PCITUS_COMB.pep.*
6: /cgn2_6/prodata/1/1aa/backfiles1.pep.*

10/2/99

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	494	11.1	398	3	US-08-931-220-5
2	494	11.1	398	5	PCT-US93-11723-5
3	494	11.1	398	5	PCT-US96-05997-1
4	140	3.2	934	4	US-08-840-466A-19
5	140	3.2	934	4	US-09-696-188B-19
6	140	3.2	10182	4	US-09-134-001C-3159
7	138	3.1	1398	1	US-08-750-532-9
8	138	3.1	1398	1	US-08-894-818B-8
9	138	3.1	1398	4	US-09-445-472-6
10	128.5	2.9	915	4	US-09-206-942-35
11	128.5	2.9	1232	4	US-09-206-942-37
12	128.5	2.9	1238	4	US-09-206-942-34
13	128	2.9	931	4	US-08-624-655A-2
14	127.5	2.9	679	4	US-08-913-942-15
15	127.5	2.9	679	4	US-09-268-347-26
16	125	2.8	741	1	US-08-277-231A-4
17	125	2.8	741	1	US-08-473-750-7
18	125	2.8	741	2	US-08-477-326-7
19	125	2.8	1168	1	US-08-620-717A-9
20	124	2.8	1341	4	US-08-296-791-3
21	124	2.8	1541	5	PCT-US95-10661A-3
22	123	2.8	621	1	US-08-328-961-2
23	123	2.8	621	1	US-08-462-397-2
24	123	2.8	1004	4	US-09-268-347-30
25	121	2.7	1167	1	US-08-485-568A-6
26	121	2.7	1167	2	US-08-590-554A-6
27	121	2.7	1167	2	US-09-184-223-6

28	120.5	2.7	992	4	US-09-206-942-61	Sequence 61, Appl
29	120.5	2.7	998	4	US-09-206-942-59	Sequence 59, Appl
30	120.5	2.7	1338	2	US-08-728-470-9	Sequence 9, Appl
31	120.5	2.7	1338	4	US-08-719-641-9	Sequence 9, Appl
32	120.5	2.7	1566	2	US-08-687-956A-23	Sequence 23, Appl
33	120.5	2.7	1599	2	US-08-617-697-9	Sequence 9, Appl
34	120	2.7	956	4	US-09-134-078-63	Sequence 63, Appl
35	120	2.7	1198	4	US-09-199-637A-405	Sequence 405, Appl
36	119	2.7	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
37	118	2.7	2411	4	US-09-268-347-36	Sequence 36, Appl
38	117.5	2.6	2137	4	US-09-134-001C-4463	Sequence 4463, Ap
39	117	2.6	1222	2	US-08-682-517-15	Sequence 15, Appl
40	117	2.6	1252	2	US-08-682-517-9	Sequence 9, Appl
41	116.5	2.6	869	4	US-08-483-101-15	Sequence 15, Appl
42	116.5	2.6	4302	4	US-09-052-469-8	Sequence 8, Appl
43	116.5	2.6	4339	4	US-09-052-469-6	Sequence 6, Appl
44	116	2.6	1336	4	US-08-551-356-6	Sequence 6, Appl
45	116	2.6	1336	5	PCT-US93-12687-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1

US-08-931-220-5

Sequence 5, Application US/08931220

Patent No. 6030835

GENERAL INFORMATION:

APPLICANT: Musser M.D., James M.

APPLICANT: Kapur M.D., Vivek

TITLE OF INVENTION: Methods and Compositions for Identifying

TITLE OF INVENTION: Group A Streptococcus

NUMBER OF SEQUENCES: 58

CORRESPONDENCE ADDRESS:

ADDRESSEE: WEIL, GOTTSAL & MANGES

STREET: 2882 Sand Hill Road, Suite 280

CITY: Menlo Park

STATE: CA

COUNTRY: USA

ZIP: 94025-7022

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/931,220

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/306,542

FILING DATE: 14-SEP-1994

APPLICATION NUMBER: US 08/160,965

FILING DATE: 02-DEC-1993

ATTORNEY/AGENT INFORMATION:

NAME: Rae-Venter Ph.D., Barbara

REGISTRATION NUMBER: 32,750

REFERENCE/DOCKET NUMBER: BAYL-004/01US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 926-6200

TELEFAX: (415) 854-3713

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 398 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: N-terminal

ORIGINAL SOURCE:

ORGANISM: Streptococcus pyogenes

STRAIN: MGAS 1719
IMMEDIATE SOURCE:
CLONE: SPEB7 (cysteine protease)
PUBLICATION INFORMATION:
AUTHORS: Kapur, V.
AUTHORS: Topouzis, S.
AUTHORS: Majesky, M. W.
AUTHORS: Li, L.-L.
AUTHORS: Hamrick, M. R.
AUTHORS: Hamill, R. J.
AUTHORS: Patel, J. M.
AUTHORS: Musser, J. M.
TITLE: A conserved Streptococcus pyogenes
TITLE: extracellular cysteine protease cleaves human
TITLE: fibronectin and degrades vitronectin
JOURNAL: Microb. Pathog.
VOLUME: 15
PAGES: 327-346
DATE: 1993
8-931-220-5

Query Match 11.1%; Score 494; DB 3; Length 398;
Best Local Similarity 30.6%; Pred. No. 7, 6e-31;
Matches 129; Conservative 73; Mismatches 155; Indels 64; Gaps 13;

QY 6 LLAIVMLFGIAMOG-----HSAPVTKERALSLARLALROVSLRMGOTAVSDKISIDYV 59
DB 10 LLSLALGFLVLANPVADQNFARNEKEAKDSAITFIQSAIAKAGARSAED-IKLDKVN 68
QY 60 ROGDAERGITSOEESGSPAYFYVANRGNNEGIALVAADDRIPITLAYSPIGRFMDSPDN 119
DB 69 LGDEL-----SGSNMYYVNIISTG--GFVIVSGDKRSPILIGYSTSGSPDANG-KEN 116
QY 120 LRMLQIY-----DOEIGLILSGKAQNLNEILRTGEGVPAEVALMDNGHFANDPMRWQ 173
DB 117 IASFMEYVQIKENKRLDTTYAGTAIEIKPV-----VKSILD-----SKGIHYNQ 162
QY 174 GYPMNKEPLLPN-----GNHAYTGCVATAAQMRYHSMPLQEGSFY----- 218
DB 163 GNPNILTPYIEKVEKQSGSFVQGHAAATGCVATATQIMKRYHNPYKGLDITYTLSSNN 222
QY 219 ----HAGSLVGNMGTGEGYMDWIN--MGNPDLNLTQSDVAVATLMDPVASVSYSEY 273
DB 223 PYRNPKNL--FAAISTROYNMNNILPTYSGRHSNVQKM--AISETLADVGISVDMDYG 277
QY 274 ENGSGTYSYVVGALRNFRYKRSLOLHVRLALYTSQEMHDMINGELASGRPVYVAGNQS 333
DB 278 PSSGASGSSRVQALAKENFQYVHNSDSKQDMEAOIKELSONPVYVYGKGV 337
QY 334 IGHAFCVDGYASDGTFFHFNMGKGVSNQFYKLLLSLTSGLIGGEGIGFTTYQELITGIE 393
DB 338 GGHAFVLDGADGRNFYHVMNGVSDGDFPRDLALNPSALGTGGAGGFGNGYSAYVGIK 397
QY 394 P 394
DB 398 P 398

RESULT 2
PCT-US95-11723-5
Sequence 5, Application PC/TUS9511723
GENERAL INFORMATION:
APPLICANT: Musser M.D., James M.
APPLICANT: Kapur M.D., Vivek
TITLE OF INVENTION: Methods and Compositions for
TITLE OF INVENTION: Identifying Streptococcus
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAW OFFICES OF BARBARA RAE-VENTER
STREET: P.O. Box 60039
CITY: Palo Alto
STATE: CA
COUNTRY: USA

ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/11723
FILING DATE: 14-SEP-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/306,542
FILING DATE: 14-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Rae-Venter Ph.D., Barbara
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: BAYL-004/030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 926-6205
TELEFAX: (415) 424-8760
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 398 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Streptococcus pyogenes
STRAIN: MGAS 1719
IMMEDIATE SOURCE:
CLONE: SPEB7 (cysteine protease)
PUBLICATION INFORMATION:
AUTHORS: Kapur, V.
AUTHORS: Topouzis, S.
AUTHORS: Majesky, M. W.
AUTHORS: Li, L.-L.
AUTHORS: Hamrick, M. R.
AUTHORS: Hamill, R. J.
AUTHORS: Patel, J. M.
AUTHORS: Musser, J. M.
TITLE: A conserved Streptococcus pyogenes
TITLE: extracellular cysteine protease cleaves human
TITLE: fibronectin and degrades vitronectin
JOURNAL: Microb. Pathog.
VOLUME: 15
PAGES: 327-346
DATE: 1993
8-931-220-5

Query Match 11.1%; Score 494; DB 5; Length 398;
Best Local Similarity 30.6%; Pred. No. 7, 6e-31;
Matches 129; Conservative 73; Mismatches 155; Indels 64; Gaps 13;

QY 6 LLAIVMLFGIAMOG-----HSAPVTKERALSLARLALROVSLRMGOTAVSDKISIDYV 59
DB 10 LLSLALGFLVLANPVADQNFARNEKEAKDSAITFIQSAIAKAGARSAED-IKLDKVN 68
QY 60 ROGDAERGITSOEESGSPAYFYVANRGNNEGIALVAADDRIPITLAYSPIGRFMDSPDN 119
DB 69 LGDEL-----SGSNMYYVNIISTG--GFVIVSGDKRSPILIGYSTSGSPDANG-KEN 116
QY 120 LRMLQIY-----DOEIGLILSGKAQNLNEILRTGEGVPAEVALMDNGHFANDPMRWQ 173
DB 117 IASFMEYVQIKENKRLDTTYAGTAIEIKPV-----VKSILD-----SKGIHYNQ 162
QY 174 GYPMNKEPLLPN-----GNHAYTGCVATAAQMRYHSMPLQEGSFY----- 218
DB 163 GNPNILTPYIEKVEKQSGSFVQGHAAATGCVATATQIMKRYHNPYKGLDITYTLSSNN 222

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:
LENGTH: 934 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-08-840-466A-19

Query Match 3.28; Score 140; DB 4; Length 934;
Best Local Similarity 17.4%; Pred. No. 0.023;
Matches 167; Conservative 128; Mismatches 331; Indels 336; Gaps 45;

74 GSPAYFYVNR--GNNEGVALVAADRIPTILAYSPIGRF-----DMSDMP-N 119
26 GLGLFFVNONSPANGENTFKLGSOSKLTHTSYO--NRLFTYTLKGTETVADLSKSDIN 83
120 L-RM---LQIYDOETIGLISGKAQINBEILRTGEPVPAEVAALNDNGHFANDPARNMG 175
84 LSTWMLNKLHLSSESEMKAPGO--QIILPLKLTPEYISAL-----124
176 PNNKEPLPNC-----NHAATGCVATAAQAIRHSHMPLOG 212
125 PLLGSAFLVAAAGVAGHINKLKMSPDVTKSNMTDDKALNAAQAASLSGQSRLNG 184
213 EGSFPIHAGSLVAGNNGSTFGGEYDWINNPGNPD-----DNLNGSOYDAATILMDRVAS 267
185 DIAKTPALG-IAGNQASS--OLOAWLQHYGAEVNLSGDNFQSSLD-----229
268 VSMSEFYEN-----GSGF-----YSYV-----284
230 FLLEFYDESEKMLAFGOVARYIDRFANLGAQREFLPANMLGYVNFIDQSGDNRL 289
285 -VGLRNFRYKRSLOLHVRLALTYSOEWMHMGELASGRPV-----YAGNNGSIG 335
290 GIGGEYWRDYFKSS-----VNGYFRMRMHESYHKRDYDERPANGFDIRNGILPSTPALG 345
336 HAFVCDGASDGTFFHFMGMGVSNGFVRLTILSPISLGIGEGGFTYIOBITTGIEPA 395
346 AKLIEQYGDNALFNSD-----KL-----GSPGAATVGVNYPPIPLVTMGI---389
396 KTPAEGTALPLALAKDIEAEKSSGLNVGYSIYNGEES--NLDLGRLNKADGE 452
390 -----DYR-----HGTGENEDLLYSQFRTQDKMSQ 417
453 VIEVKTSSINISWYGEHPESFSLAPNLSGINTITLLYR-----TGTQ 500
418 QIEPQ-----YVNELRLTSSGRDLYORNNITILEYKKODILSLNIPHDINGTEH 467
501 -----W-EPVHAAGGVNLSIKVNTTDPNNVVYVVDNNEKLSIVP 540
468 STOKIQLYKSKYGDRIYWDMSALRSQGOIHSQSADQYAILPAYOGS-----522
541 NSFVADLNSYHS-----TIVQNSDSPDIRTPAFAIATGATADVLSLGMW 591
533 NIYKVTARAYDRNGSSNNVOLLITTVLSNGOVYDQVGTFTADKTSAKADNDADITTYA 582
592 AEVPGG--SSNYPVVMSKDVLTJLSEGGYTLMYRESINNOKEDEKKIGSVKPTPEYTP 649
583 TVKNGVAQAQANVPVPSFN-----IYSGTATL-----GANSKATDANGAKATYLLSSIP-629
650 LFEVGHNOTSTYTLIDMAHNRVLPDFLKNLGLPFNELVYVROTOSSSGSLMAAQETVH 709
630 -----GQVYVSAKTAEMS-----ALNSAVIIFDQTKASITEI-KADKTTA 670
710 IKOGE--TFVYKPVYEG-PIPGSTR-ATLAAVNGQOOLYL-----K 748
671 VANGKDAIKTVYKVMNGQPVNNOSYVTFSTNGFMNGSKSQQTATGNGRATITILTSSA 730

QY 749 GRNRYTVNGTAVEATESS--EIRY-----FPNARDYVEISAPC--- 789
DB 721 GKATVSATVSDGAEVAKATETFFDELKIDNKVDIGNNVNGELPNITLQYGFRLKASG 790
QY 790 -----IPQETSLFPLSKRYKVNLSAGHRMDVSRPLNG-----AYILKVGCTTK 838
DB 791 DGTYSWTSSENTSIATVDASGRVTLN-----GKGSV-VIKATSGDKQVTSYTIKAPSYM 844
QY 839 IN 840
DB 845 VD 846

RESULT 5

US-09-696-188B-19
Sequence 19, Application US/09696188B
Patent No. 6406885

GENERAL INFORMATION:

APPLICANT: Stewart, C. Neal
McKee, Marian L.
O'Brien, Alison D.
Wachtel, Marian R.
TITLE OF INVENTION: Method Of Stimulating An Immune Response
By Administration Of Host Organisms That Express Intimin
Antigens.
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flinnegan, Henderson, Farabow, Garrett &
Dunnet, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/696,188B
FILING DATE: 26-Oct-2000
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/840,466
FILING DATE: 1997-04-18
ATTORNEY/AGENT INFORMATION:
NAME: Boone, Laura L.
REGISTRATION NUMBER: 43,505
REFERENCE/DOCKET NUMBER: 04995, 0029-00000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 934 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-696-188B-19

Query Match 3.28; Score 140; DB 4; Length 934;
Best Local Similarity 17.4%; Pred. No. 0.023;
Matches 167; Conservative 128; Mismatches 331; Indels 336; Gaps 45;

QY 74 GSPAYFYVNR--GNNEGVALVAADRIPTILAYSPIGRF-----DMSDMP-N 119
DB 26 GLGLFFVNONSPANGENTFKLGSOSKLTHTSYO--NRLFTYTLKGTETVADLSKSDIN 83
QY 120 L-RM---LQIYDOETIGLISGKAQINBEILRTGEPVPAEVAALNDNGHFANDPARNMG 175

```

Db      84  LSTIMSLINKLYSESEEMKAAFGQ--QITLPLKLPFEVSAL-----124
Qy      176 PMNKEPILPNC-----NHATYGCVATAAQMRYHSPLOG 212
Db      125 PLIGSAPLVAAGVAGHTNKLTKMSPDYKSNMTDRLALNYAAQOASLSQSLSLNG 184
Qy      213 EGSFDYHAGSLVNMGCTGEMTWMPCNPDL-----DNLQSOVDAYATLMRDVSAS 267
Db      185 DYAKDTRALG-IAGNOASS--QLOAMLOHGTAEVNIQSGDNFEGSSLD-----229
Qy      268 VMSKFEYEN-----GSGT-----YSVYV-----284
Db      230 FLFPYDSEKMLAFGVGARYIDSRFTANLAGOQREFLLPAMMLGVNFIQDPSGDNTRL 289
Qy      285 -VQALNNRRYKSLDHLRYALCTSOEHDHMGELASGRPVY-----YAGNNOSIG 335
Db      290 GIGGEYWRDYFKSS--VNGYFRMRMRHESYHKKDYDERPANGFDIREFNGYLPSPYALG 345
Qy      336 HAFVCDGVASDCTGFHFMWCGVSGNGFYKLTLSPTSLGIGEGIFTIYQETITGIEPA 395
Db      346 AKIYQIYGDVNLAFNSD-----KL-----QSNPGAIYGVNTPITPLTYMGI---389
Qy      396 KTPAEAGTALPLALKIEAEYKSSGLNVGYSTYNGEES--NLDGYRLNRADGE 452
Db      390 -----DTR-----HGTGNEMLDLYSMQFRYQFDKSMQ 417
Qy      453 VIEVKTSINISMYGGEHPESLAPNLSOGINFTLLYR-----TGTG 500
Db      418 QLEPQ-----VYNELRTLSGRYDLVORNNNILEYKKODILSLNIPHDINTGE 467
Qy      501 -----W-EFVRAAGGVNSIKVNTDPNNVVVTVNNEKSLIYV 540
Db      468 STOKIOLIVKSKYGLDRIMWDSALRSQSGQIOHSSQSADYQALIPAYOGS-----522
Qy      541 NSFVADLNSTEHS-----TIVQFNSDSPDEIRTPVAFALSTGATADVYISLGWM 591
Db      523 NIYKTVARAYDRNGSNNSNNVQLITIVLSNGOVVDQVGTDTAKTSAKADNADTIYTA 582
Qy      592 AEPGCG--SSNPVYVSKDVLTLSEGDYTLMYRFSINNOKDEKKIGSVSKPTETHP 649
Db      583 TYAKNGVQAQNVVSNF-----IVSGTATL-----GANSATDANGKATVILKSTP---629
Qy      650 LPEVGHNOTSTYTLDMAHNRVLDPFTLKNLGLPFNGELVYVFRQTOSSGSLMAOETVH 709
Db      630 ---GGVVVSAKAEHSS-----ALNNSAVIFFDOTAKAITEI-KADKETA 670
Qy      710 IKOGC--TFVYKPVVEG-PIPDGSYR-ATLHAFVNGQOQLYL-----K 748
Db      671 VANGKDAIKYTVVMKNGQPVNNQSVTFSTFGFMFGKSOQTATTCGDRATITLSSSA 730
Qy      749 GRKNYVYKIVNGFAVEAIES--EETRV-----FPNARDVVEISAPC---789
Db      731 GRATVATVSDGEVATEVTEFDELKIDNKVDIGANNVAGELPNWLOGQEFKLASGG 790
Qy      790 -----IPESTIILDLGKIYVKNLSAGHGRMVSRLPNC-----AYLLVYDGYTK 838
Db      791 DGTYSMYSENTSIATYDASGKVTLN-----GKGSV-VIKATSDKQTVSTIKAPSMIK 844
Qy      839 IN 840
Db      845 VD 846

```

```

RESULT 6
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007

```

```

; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match      3.28; Score 140; DB 4; Length 10182;
Best Local Similarity 18.88; Pred No. 1;
Matches 149; Conservative 116; Mismatches 282; Indels 244; Gaps 42;

Qy      148  TEGVPAEVALMDNGHEPANDPMRWNOGYPMNNEPILLPNCNHAITYGCVATAAQMRYHS 207
Db      1819  SQSIPKQIYETTINGRP-----NSSGDVYPCNFVQAVNOY-----1853
Qy      208  WP-----LQSGSPDYHAGSLVGNMSTGFGEY-----DWIM-----PCNPDLDN--248
Db      1854  WPEHMDFRMAQSGGT--PSSRNAGSETKTVYVYONGQTEYNNVLFKVPKMRPVIDSNS 1910
Qy      249  -LQSOVDAYATLMRDVSASVMSFEY-NG-----SGTYSVYVVAL-RNNFR 293
Db      1911  VISKQNGOQIILYRNPQAOVTLVOSKTVLPNTWTIDNSGATVTTQGLPLPGNIT 1970
Qy      294  YKRSLOLHYRALYTSOEHMIMGELASGRPVYAGNNO--SIGAFVCDG-----342
Db      1971  AKTSMTNNV--TYTKQN--SSGIASNTTEDISFSENDDQVNVYAGQAKMDGIKIKGTN 2027
Qy      343  -----YASDGFHFMWCGVSGNGFYK-----LTLSPSLGIGEGIFTIYQ 386
Db      2028  YNENDFNSFISNIPASHTLTWNEEPNSWKNNIGTTTKTVTVTLPHNQGTRFVDIPITIP 2087
Qy      387  EITGTIEPAKTPAE-----AGTDALPLT-----ALKDIEAEYKSEGLN- 425
Db      2088  TV-----TAKNPYRQDKGRMLTNGTGYNNITIEENNRRLGTSKMDNRDPDKNIAGVON 2142
Qy      426  ---VGYSIYNGEES--NLDGYRLNRADGEVIEKTSINISMYGGEHPESF 475
Db      2143  LIALVNPPIGSTPLEVYVYVYVNFDPQIYK-----IQVGTFFPKGTAGYKHLDEG 2197
Qy      476  SLAP-----NQLSGITITILLYRRTGEQEPYRHAQGVNSIKYNTDDPNVYV 527
Db      2198  EGPLIDGKPEYMNQOSTG-----TTSQMOQLAYTRTFPVKGTGYDVNPSMWGV 2247
Qy      528  TVDNNCKLSIVNSFVADLNSYEHSHITVQFNSDSPDEIRTPVAFALSTGATADVYSL 587
Db      2248  WQTSQSAKF-IVYNA-----KRNQPIIT--OSKTDVYTP-----GAYRNLILIS- 2289
Qy      588  GWMAVEYGGSSNPVYVSKDVLTLSEGDYTLMYRFSINNOKDEKKIGSVSKPTETHT 647
Db      2290  -----GTNXY-IOASADKIVINKNGKL--TFVYKANNDRW-----IVENGSP---2329
Qy      648  HPLFVGHNOTSTYTLDMAHNRVLDPFTLKNLGLPFNGELVYVFRQTOSSGSLMAOET 707
Db      2330  -DINGIGPTNNGT-AISLSRLAVRPGSISALATGSGEETI-----STATSELY-----2377
Qy      708  VHIKOGTEFYKPVVGEPIPDGSYRATLHAFVNGQOQLYLKGR-----NYTVK 756
Db      2378  -----YKAPQPE--OATHTYDNGTFDILPDNSRNSLNPTEREVINTTEK 2421
Qy      757  IVNGTAVE--AIESSEIRVFPNPARDYVEISAPCIPESTIILDLGKIYVKNLS- 811
Db      2422  L-NGNEFQSKFTITKNNNGKWTINKNPNVYEFN-----QNGKVVYF--SANTIKPNSQIT 2473
Qy      812  ---SAGHGRMD 819
Db      2474  ITPKAGGNGTE 2484

```

RESULT 7

US-08-750-532-9

Sequence 9, Application US/08750532
Patent No. 5736339

GENERAL INFORMATION:

APPLICANT: MITTA, Masanori
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MORISHITA, Mio
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENE
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
STREET: 419 Seventh Street N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/750,532

FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP95/01095
FILING DATE: 05-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 1994/130236
FILING DATE: 13-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 1994/173912
FILING DATE: 26-JUL-1994

ATTORNEY/AGENT INFORMATION:

NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: MITTA-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528

INFORMATION FOR SEQ. ID NO: 9:

SEQUENCE CHARACTERISTICS:
LENGTH: 1398 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: peptide
US-08-750-532-9

Query Match

Best Local Similarity 3.1%; Score 138; DB 1; Length 1398;
Matches 157; Conservative 91; Mismatches 262; Indels 296; Gaps 41;

QY 185 PGNHAYTGVATAAQAQIMRYHSMPLQEGSFYDHASGLVNGMGTGEMVDMNPNP 244
DB 351 PNCGEAVFG-----WDGHHGT--HVAAGTVAGYDSN--NDAMDMLSMYSGE 392
QY 245 -----DLNLTQ-----SQVDAYATIMRDVSAS----- 267
DB 393 WEVFSRLYGMVDTVTDTVGVAQAQIMAIRVLRSDGSGMMDIEGMYAATFGADV 452
QY 268 VSMSEF-----YENSGTYVYV-----VG 286
DB 453 ISMGLGNAPYLLDGTDESVAAVELTEKYGVFVIAAGNPGINIVSGVATKAITVG 512
QY 287 ALRNFYKRSIQLHVALYTSQEMHDMINGELASGRPVY-----AGNDSIGHAFC 340

DB

513 A-----AAVPIV--GVYVSO-----ALGVPDYGGFYPPATVNR/IAF-ESS 552

QY 341 DGVASDGTFFHFNW---GMGVSNGEYKLLSPSTLSIGGEGGFTIYETITGIEPAKT 397

DB 553 RGRPRIDEIKPNVAPRG-----YSSLPMWIGG-----ADFGSGTSM-T 593

QY 398 PARGDADPLTLAKDIEAEYKS-----ESGLNVGSIYVNGEESNLDLGYR----- 445

DB 594 PHVSGVALLISGAKKEGIYVNDILIKYLESAGTWLEDDPYTGKYTELDGGRHVLVNT 653

QY 446 -----LNKADGEVIEYKTSINISWYCGEHESESFLAPNOSGINITTLTYRTGTE- 499

DB 654 KSWELIKALNGTTLPIYDHMAKSYDFAEY-----LGVDVIRGLYARNSPD 701

QY 500 --QWEVYRAAGGYVNSIVNTDP-----NNVYVVDNNEGKLSIVPNSFVADLNSYE 551

DB 702 IYEMH-IKYVGDTEYKFEIYATEPMIKPFVGSVILENTEFVLRY-----KYDGELE 755

QY 552 HSTIYVQFNSDSPDEIRFPVAFALSTGATADVISLGWVMAEVRGSSNYPVMSKDVLT 611

DB 756 PGLYGRITIIDP-----TTPV-----IEDEL-----NTIYIEPK---F 787

QY 612 LSEGYTL-WRFPSINOKDEMKKIGSVYKTPLEYTHPLFEVGHNOTSTYTLDMANNRV 670

DB 788 TPENNYTILTWY-----DINGPEVTHHFTTYVEGVYLAAMTYWDYG 830

QY 671 L--PD-----FTLKNL-----GLPFGELVYV---FROTOS-----SGSLW 702

DB 831 LYRPGMVFVQYLDYLAANSNPPGMEWLGTFENFADLYEGFLVRIYGEIETPSVW 890

QY 703 AAQET-----VHIKQGEYFVKPVYEG--PIPDGSYRATLHAIVNGOOLKGRNY 753

DB 891 YINRTYLDNTNEFSIEFNTINIVAPINATLIPILGIVNMSVESV--GDEFFIKG--- 944

QY 754 TKIYNGTAVPAISSSEIRFPRPADYVEISAPCI-PQETSIILFDLSGKIYMNLS 812

DB 945 -----IEVPGTIELK-----IRIGNPSVNSDLIDLYLDSKGNLVALDGNP 986

QY 813 AGHRMDVSRIPNGAYILKYDGYTK 838

DB 987 TAEEVVEYKPKGVYSIVVGYSVR 1012

RESULT 8

US-08-894-818B-8

Sequence 8, Application US/08894818B
Patent No. 6261822

GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435

Db 719 TGSVEVTAKTGDIKGIIESNSGVNITASGDTLNV 754

RESULT 12

US-09-206-942-34
Sequence 34, Application US/09206942
Patent No. 6432669
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M.
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H.
TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae H1gH
TITLE OF INVENTION: Molecular Weight Proteins
FILE REFERENCE: 1038-861 M1S:Jb
CURRENT APPLICATION NUMBER: US/09/206,942
EARLIER FILING DATE: 1998-12-08
PRIOR APPLICATION NUMBER: 09/167,568
EARLIER FILING DATE: 1998-10-07
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 34
LENGTH: 1228
TYPE: PRF
ORGANISM: Haemophilus influenzae
US-09-206-942-34

Query Match 2.9%; Score 128.5; DB 4; Length 1228;
Best Local Similarity 19.8%; Pred. No. 0.29;
Matches 102; Conservative 79; Mismatches 220; Indels 115; Gaps 20;

QY 129 OEGLIL-----SGKAOLNEILRTGVPAAVHA-----LMNGHFANDPMRNOG 174
DB 329 QYTVIEEONSASGSSIK---FKSEG---STHAFITKNDLILANGGNTSLNOVAG 382
QY 175 YPMNNEPPLPNGNHAHYTCVATAAOLMKYHSMPLQEGSPDYHAGSLVNGSGTGGEM 234
DB 383 IDSNLKSLIANKNITFEAGNITLAD-----KKPIKGNITVKEGANTLRSANYGND 437
QY 235 YDMINRPN-PDLNITOS-----QVDAVATLMDVSVASVMS--FYENGSGTYS 281
DB 438 KSNLSIRGNVTNKNLTVGSALINIEKNLTVEGSAFLANPNYSFVSGLEFNOGKSNIS 497
QY 282 VYVVGALRNFRKRSLOLHVRLYTSQEWHDIMRGELASGRPVYVAGNNOISGHAFCVCD 341
DB 498 IAKGGLIFKDIETGSLNITTK---SDSNHHTIIRKGNITRKCDLITNN-----544
QY 342 GYASDGFPHNMGVNGSVNGFYKLLLSPTSLGIGEGIGFTIYOELITGIEPAKTPAEA 401
DB 545 ---GDWT-ELQIG-GNISQKEGNTL-----ISSDKVNIETRIITIKAGVNGDSDSNE 591
QY 402 GTDALPILAKDIEATKESGSLNVCYSIYN---TGEOSNLDLGRLLKADGEVIEVK 457
DB 592 AYSANLITTK---ELKLTNDLNI---SGFNKAETIAKONSNTLIDNSAGWTDAKKYT 645
QY 458 TSSINISWYGEHPESFSLAPNOLSGITITLRLRTGTEQMEVVRHAGGVASIKV 517
DB 646 FSNVKS-----KISASHNVTL-----NS-KV 667
QY 518 NTDPPNNVYVDNNEKSLIYVNSFVADLNSYHSHITVVOFNSDSDPELIRTPAVALST 577
DB 668 ETSGDPTDSTEDGNNNTGLITFAKNVTYNNNITLSHKTVNITASENVTYKAGTTIN--AT 724
QY 578 GATADVDYISLGMVAEVPGGSSNYPVWSKDVLTLS 613
DB 725 TGSVEVTAKTGDIKGIIESNSGVNITASGDTLNV 760

RESULT 13
US-08-624-655A-2
Sequence 2, Application US/08624655A
Patent No. 6323005
GENERAL INFORMATION:

APPLICANT: DABAN, MONTERRAT
APPLICANT: MEDRANO, ANDRES
APPLICANT: ESPUNA, ENRIC
APPLICANT: QUEROL, ENRIQUE
TITLE OF INVENTION: TRANSFERIN-BINDING PROTEIN 1 (TBP1) GENE OF
TITLE OF INVENTION: Actinobacillus pleuropneumoniae, ITS USE TO PREPARE
TITLE OF INVENTION: PRODUCTS FOR THE UTILIZATION IN VACCINES FOR
TITLE OF INVENTION: PLEUROPERNEUMONIA AND AS DIAGNOSTIC REAGENTS
FILE REFERENCE: P00740
CURRENT APPLICATION NUMBER: US/08/624,655A
CURRENT FILING DATE: 1996-03-22
PRIOR APPLICATION NUMBER: 95 00592
PRIOR FILING DATE: 1995-03-24
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 931
TYPE: PRF
ORGANISM: Actinobacillus pleuropneumoniae
US-08-624-655A-2

Query Match 2.9%; Score 128; DB 4; Length 931;
Best Local Similarity 19.3%; Pred. No. 0.2;
Matches 146; Conservative 104; Mismatches 286; Indels 222; Gaps 43;

QY 57 YVTRQG-----DAE---RGIT-----SOEGSPATFYVANNRNNEGYALVAADRPTI 102
DB 208 YTHRDKETRAHKAHDAESRSONITRVGETNELDTSNRYTTNNQHTYGMFLIKDEPPL 267
QY 103 -LAYSPIGRFMD-----SWPDLRMWLOIYDOEIGLISGAOLNEILRTGVPAAVHA 156
DB 268 DCFPKAMARTKTPSPRSYETTPREKQAYENQKH-----TERLNADYIGEYR 318
QY 157 ALMDNGHFANDPMRNOGYPMNNEPPLPNGNHA-----YTCVATAAOLMR---204
DB 319 ALDPLKYSDSMLVLYGTFSEKHYAGYEHSEKORYDRTDYTAWPOPSDLRTGRN 378
QY 205 -----YHSMPLQEGSPDYHA-----GSLVGNM--SGTGEYMDWI 238
DB 379 WYPMNNAKGLYRNDALDGV-AIDFTEDGYKSSKGLMKAARIDEMHTDRGLALRYT 437
QY 239 NMDGNPDLNLT-----QSOVDAYATLMDVSVASV-----SMSFE 274
DB 438 NQGNRLIDRLSLSPDQKINLSTRLKNNCEPTIDKRCRATLRLKMSSTKNEOSYE 497
QY 275 NGSCTYSV-----YVVGALRNFRKRSLOLH-VRLYTSQEWHDIMRGELAS 321
DB 498 EKHDTIQLSDKTVOTGLGKHQNLMLGSDRFNSTLRHRLISEFSVGTWH-RIRNGYK 556
QY 322 GRPVYVAGNNOISGHAFCVCD-----G 342
DB 557 DTPYITELKQALYSKNECYSCTIAGRADCATSKIGHNHYIALRDNFAITYLIDIGL 616
QY 343 YASD-STFHRNMW---GVSNGFYKLLLS-PTS-LGI---GEGIGFTIYOELI---389
DB 617 YRDKDKFRSTHMANAGDVKNSAMNIGIYAKTSPSLSTRASSGGRVSPFQELRLRY 676
QY 390 TGIAPAKTPAEACTDAL-PILAL-KDIEATKESGSLNVCYSIYNNGEOSNLDLGRRLN 447
DB 677 DGAHKSSDAYQKTEKISPEKSLNOEVAATFKDDFGV-VEVSFKNDYKOL-TAPERMH 734
QY 448 KAGG-----EIVFKNLSIN-----ISWYGEHPESFSLAPNOLSGITITLRLRTGT 498
DB 735 QIOSMNIYFNVDIKLDGILIKLDMNGVFD-----KIPBITY-TLAYSIMRY 783
QY 499 EOWEPYRHAOGGYVN-SIKVNTTDPNNVVVTV--DNNEKSLIYVNSFVADLNSYHSH 554
DB 784 KE---VKNYQ-GYMNIRSPDLDTIOPARYVVGIGYDQDEKMGV-----N 824
QY 555 ITVOENS-DSPDEIR--TPYAFALSTGATADVDYISLGMVAEVPGGSSNYPVWSKDVLT 611
DB 825 LVTHTSSGKNPNELRGNNEGVGFA-NYERTATKRTKRLSMHTFDLTG---YITPMKH--TT 877

QY 612 LSEGDYTLV-VAFSINNOKDEMKKIGSVKPTFTEYTH 648
DB 878 VRAGVYNLMNRYT-----TWESYROSSLAINHOTH 909

RESULT 14
US-08-913-942-15

Sequence 15, Application US/08913942
Patent No. 6200578
GENERAL INFORMATION:
APPLICANT: St. Geme, Joseph
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Flehr Hohbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,942
FILING DATE: 29-DEC-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/409,995
FILING DATE: 24-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/4031
FILING DATE: 22-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vance, Dolly A.
REGISTRATION NUMBER: 39,054
REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAY
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 679 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-913-942-15

Query Match 2.9%; Score 127.5; DB 4; Length 679;
Best Local Similarity 19.5%; Pred. No. 0.13;
Matches 138; Conservative 89; Mismatches 239; Indels 241; Gaps 33;

QY 242 GNPDLNL-TOSQVDAYATLMRDVSASVSMSEFYENGSGTYYVVGALRNFRKRSIOL 300
DB 86 GTTEVINLMTDSSGNAVGSSTTFKAGDNLKIKOSG-----NDFTY--SLK 130
QY 301 HVRALYTSQEHMIMINGELASGRPRYYAGNNSIGHAFVCDGYASDGTFFHFMGMGVSN 360
DB 131 ELKNL-TSVETEKLSFG--ANGNKVDITSDANGKLAKTGNGNGNSVHLN---GIAS 183
QY 361 -----GEYKL-----TLSPISLIGGEGIGFTTYOETITGIEP 394
DB 184 TLTDLAGGTGHTVDNIDAVNTHRAASVQDVLNSGNNIGGNNVDVFRYYDVFVNG 243
QY 395 A-----KTPAEGTALPILALDLAEYKSESQ--LNVGSIYMTGGEOS 438
DB 244 ANANVSATADTAHKKTTVRVDVYGLPV-----QYTEDGKTVYKVGNEYK-ANDDG 294

QY 439 NLDGLYRLNKADGEVIEVTSINISIMYGGEHPESFSLAPNLSOGINTITLLYRRGT 498
DB 295 SADMNQKVE--NELAKTKYKLVAS-----GTPNVRKISNADGT 332
QY 499 EOMEPEV-----RHAOGGVYV-SIKVNTDPNNVVYTVVNNEGKLSI 538
DB 333 EDTDAVSFNOLKALODKQVTLSTSNAYANGGTDNCGKATQTLNSGLNFKFSSGELL 392
QY 539 VPNSFVADINSYEHSTITVOFNSDSPDEIRTVARALSGAL-----ADVLSL 587
DB 393 I--SATGDIVTTPPKGSIVQVGGDKAST-----SKANTTEGLVEASELSLKL 442
QY 588 GW-VMAEVPG-----GSSNYPVWSKDYLLTSEGD-----YTLVYRFSINNOK 629
DB 443 GKVGVGEKVGSELDGTSKETLVKSGDKVTLAGDNLKVKQEGTFYAL-----K 493
QY 630 DEWKKIGSVYKTPPEYTHPLFEVGHNGTSTYTLDMANRVLDPFTLLKGLPENGELV 689
DB 494 DELTGKVSVEFDYAN-----GANAST-----KITKDLTTLTLANGANATVY 537
QY 690 VERQTOSSGSLMAAQETVHVKQGETFYKPYEGPIPDGYSRATLHAFVNGQOQV--- 746
DB 538 DADKIKVASDGISAGNKAV-----KNVAAGEI-----SATSDAING-SQLAYVA 581
QY 747 -----LKKRNYTVKIVNGTAVEAIESESELRFPNPDARDVEISADCIPOET 794
DB 582 KQVTLACQVNNLEKGVKVRKADAGTASALASQ-----LPQAT 622
QY 795 STIIFDLGKIVYKNSLSAGHR-----MDVSRLL-PNGAYILKVDGYT 836
DB 623 -----MPGKSWVSIASSSYOGQNGLAIGVSRISDNKVIIRLGGTT 663

RESULT 15
US-09-268-347-26

Sequence 26, Application US/09268347
Patent No. 6335182
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M.
TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
FILE REFERENCE: 1038-860
CURRENT APPLICATION NUMBER: US/09/268,347
CURRENT FILING DATE: 1999-03-16
NUMBER OF SEQ ID NOS: 54
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 26
LENGTH: 679
TYPE: PRP
ORGANISM: Haemophilus influenzae
US-09-268-347-26

Query Match 2.9%; Score 127.5; DB 4; Length 679;
Best Local Similarity 19.5%; Pred. No. 0.13;
Matches 138; Conservative 89; Mismatches 239; Indels 241; Gaps 33;

QY 242 GNPDLNL-TOSQVDAYATLMRDVSASVSMSEFYENGSGTYYVVGALRNFRKRSIOL 300
DB 86 GTTEVINLMTDSSGNAVGSSTTFKAGDNLKIKOSG-----NDFTY--SLK 130
QY 301 HVRALYTSQEHMIMINGELASGRPRYYAGNNSIGHAFVCDGYASDGTFFHFMGMGVSN 360
DB 131 ELKNL-TSVETEKLSFG--ANGNKVDITSDANGKLAKTGNGNGNSVHLN---GIAS 183
QY 361 -----GEYKL-----TLSPISLIGGEGIGFTTYOETITGIEP 394
DB 184 TLTDLAGGTGHTVDNIDAVNTHRAASVQDVLNSGNNIGGNNVDVFRYYDVFVNG 243
QY 395 A-----KTPAEGTALPILALDLAEYKSESQ--LNVGSIYMTGGEOS 438
DB 244 ANANVSATADTAHKKTTVRVDVYGLPV-----QYTEDGKTVYKVGNEYK-ANDDG 294
QY 439 NDLGLYRLNKADGEVIEVTSINISIMYGGEHPESFSLAPNLSOGINTITLLYRRGT 498

Db 295 SADNOKVE--NGELAKTKVLSAS-----GTNPVKISNVADGT 332
QY 499 EQMEPV-----RHAOGGYV-N-SIKVTPDPNNVYVVDNNEGLST 538
Db 333 EBDVAVSFKQALKQKQVLTSTSMAYANGTNDGKAKQTLISNGLNFKKSSDGLK 392
QY 539 VPNSFVADLINSYESTITVOFNSDSPDEIRTPVAFALSTGAT-----ADYISL 587
Db 393 I--SANGDYTFPPKGSVOVGDDGKASI-----SKGANTTEGLVEASELVESEINKL 442
QY 588 GW-VMAEVRG-----GSSNPVYVSKDVLTLSEGD-----YTLMYRFSINNOK 629
Db 443 GMYVEKYGSGELDGTSTKTLVSGDKVTLKAGDNLKVGQEGTNFYAL-----K 493
QY 630 DEKKKIGSVYKPTETHPLEFGVHNGTSTYTLDAHNRVLPDFTLKNGLEFPNGELVY 689
Db 494 DELTGKSVYEFKQTAN-----GANGAST-----KITKDGTLITLANGANGATVY 537
QY 690 VFRQTOSSGSLMAAETVHIKGETFVYKPYVEGPIDGSRATLHAFVNGQOOLY--- 746
Db 538 DADKIVASDGIAGKAV-----KNVAGEL-----SATSDAING-SQLYAVA 581
QY 747 -----LKGRNTYVTKIVNGTAVAEISSEIRVFPNPARDYVEISAPCIPOET 794
Db 582 KGVTLNLAGOVNLEGRKYNKVKRADAGTASALAAQ-----LPQAT 622
QY 795 SILFLSGKIYVKNLSAGHR-----MVSRL-PNGATYILKVDGYT 836
Db 623 -----MPGKSMVSIAGSSYOGNGLAIGVSRISDNGKVIIRLSGTT 663

RESULT 16

US-08-277-231A-4
Sequence 4, Application US/08277231A
Patent No. 5643725

GENERAL INFORMATION:
APPLICANT: Green, Bruce A.
APPLICANT: Brinton, Charles C.
TITLE OF INVENTION: Sequence and Analysis of LKP Pillin
Patent No. 5643725
TITLE OF INVENTION: Structural Genes and The LKP Pill Operon of No. 5643725cytable
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/277,231A
FILING DATE: 19-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Alice O.
REGISTRATION NUMBER: 32,542
REFERENCE/DOCKET NUMBER: ACC94-02
TELEPHONE: (617) 861-6240
TELEPHONE: (617) 861-6240
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 741 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-277-231A-4

Query Match 2.88; Score 125; DB 1; Length 741;
Best Local Similarity 21.18; Pred. No. 0.24;
Matches 128; Conservative 76; Mismatches 216; Indels 188; Gaps 32;

QY 22 APTKREKALSLARLAR-----QVSLRNGQTAVSQKISIDYVROGDAERGITSQEEG 74
Db 186 APYVRCANNTNAVSIKONQYITQITVPAGPVIINDLYASG---SGDLTVEL-QESDG 241
QY 75 SPAYFYVANGNNEGVALVADRIPTILAYSPIGRFDMSMDNLRMWLOIYDQET- 132
Db 242 KVSFIVP-----FSNLAPLRVGLHRLQAGGRYRIDS-----RTFDERVLOG 285
QY 133 LI--LSGRQQLMBELLRTGCVPAEVAHALDNGHFADPM-RNNQGVPMNKE-PILPNG 187
Db 286 VIOGYGLNHLTLNLSLTYTHRYAGLE---GFLGNTPIGAFSADATWSHAEPPLKHS 340
QY 188 NHAY-----TCGVATAAQIMRYHSMPL----- 210
Db 341 KNGYSLHGSYSINPNESGNTITLAA--YRISSDFTYLSDTIGLNRTPROSCATYLPET 398
QY 211 QGEGSEFDYHAGSLVGNKSGTF--GEMYDWIMPGNDPLDNLTOQVDAVATLMDVASY 268
Db 399 RPKNQOVSLQSLGNNGNLYLSGQYVNYWEKRGTN-----TOYOV-AVNSFHLNYSV 452
QY 269 SMS--FYENGSGTYSYVVGAL-----RNNFRYKS-----LOLHYRLYTSQEHDM 314
Db 453 NLSQSIDKETGKRDNSITLISLPLDGNHSADSSYSGNDINORLGVNSF--GERHGW 510
QY 315 IRGELASGRPVYVYAGNQSIGHAFVCDGYAS--DGFHFNGMGVSNGFYKLTLLSPTSL 373
Db 511 SYGINAS-----RNNQ-----GYISYGNLSHNSISGY--RAYSISROSLKRSI 553
QY 374 GIGEGE-----GFTIYQETITGIEPAKTPAENGDAIPALAKDIEATKSEGLANG 427
Db 554 SLGASGAVVAHKRGITLSQPV-----GESFAIHAKD--AAAKVSESGANVS 598
QY 428 YSIYNGEBSNLDLGRLUKADGEVLEVNTSSINISWYV-----GEHPSFSLAPQLS 483
Db 599 LDYF-----GNVAMPYTSPEYINWITGINSDAEANYFEATERQIT 639
QY 484 QGINTTILYLRRTGEQMEPVRRHAGGYVNSIKVNTDPNNVY---TYDNNEGKLSIV 539
Db 640 PRANSISLVDFRIGN-----TMVLENTLPLNGEPYPMASATQDSEG----- 681
QY 540 PMSFVADL 547
Db 682 --AFVGDV 687

RESULT 17

US-08-473-750-7
Sequence 7, Application US/08473750
Patent No. 5834187
Patent No. 5834187 5786143
GENERAL INFORMATION:
APPLICANT: Green, Bruce A.
APPLICANT: Brinton, Jr., Charles C.
TITLE OF INVENTION: Sequence and Analysis of LKP Pillin
Patent No. 5834187 5786143
TITLE OF INVENTION: Structural Gene and the LKP Pill Operon of No. 5834187 5786
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: US
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

Query Match	2.88;	Score 125;	DB 2;	Length 741;
Best Local Similarity	21.18;	Pred. No. 0.24;		
Matches 128;	Conservative 76;	Mismatches 216;	Indels 188;	Gaps 32

```

OY 22 APVTERALSLRLR-----OVSJLMGOTAVSDKSIDYVRROGDAENGITSOEBG 74
Db 166 APVYGVANTNKAKVSIKONGTITTYITVPAGFVINDLASY-----SGLVETI--QESDG 241
OY 75 SPATYYVANRKNNEGTALVADDRIPITLAVSPIGRFWDSPMDLRMWLOIYDOEI--G 132
Db 242 KYRSEIVP-----FSMLAPLRMGHRLYOLAGRRYIDS-----RTERDEVLOG 285
OY 133 LI---LSGKAOLNEELRTGEGVAPVHMLMDGHPMP--RMNGYPMNKE--PLLENG 187
Db 286 VLOYGLTNHLTNSLSLTRYHRAGLF-----GELNPIGAFSADATWNAHEFLKHYV 340
OY 188 NHAY-----TGCVATAAQIMRYHSWPL----- 210
Db 341 KNGYSLHSGYSTINFESEGTNITLLA--YRYSRDEYTLSDITGLNRTRPFGSAYLPEIT 398
OY 211 QGEGSFYDHAAGSLVGNWMSGTF--GEMYDWMIMPGRDLDNTQSOVDAYATLMRVASV 268
Db 399 RPKNOFOYSLQSOSLGMGNLTLSGOTYMYEKRGTN-----TOYQV--AYSNPHILANSV 452
OY 269 SMS---FEVNSGTYSVYVVGAL-----RNNFRKRS-----LOLHRALEYSEWMD 314
Db 453 NLSOSIDKEKTRKRONSTIYLSLPLGDNHSDSSTSRSGNDINORLGVNGSF--GERHOW 510
OY 315 IRGELASGRPVYVAGNNOISGHAFVCDGYAS--DGTFRHNMWGGVSNFYRLTLLSPSTL 373
Db 511 SYGINAS-----RNNQ-----GYRSYDGNLSHNNSIGSY--RASYSROSLKRNISI 553
OY 374 GIGEGEI-----GFTYOEITIGIEPAKPRALAGTALPLAKDIEAEKSSGLVNG 427
Db 554 SLGASGAVNAHKHHTLSQPY-----GESFALIHAKD--AAGAKVYSGANVS 598
OY 428 YSIYNGEEOGNDLGYRLKKAODEVIEVKTSSINISWGY---GEHPESFSLAPNOLS 483
Db 599 LDYF-----GNAVMPYTPSPEINYIGINPESDAEVAEEFATERQII 639
OY 484 QGINTITLLRRTGEQEWEPVRAHAGGVYVSIKNTDPPNNVY---TYDNNGRKSTIV 539
Db 640 PRANSISLVPRCKN-----IMVLFNLTLPMGEPRVPMASTAODSEG----- 681
OY 540 PNSFVADL 547
Db 682 --AFVGDV 687

```

Sequence /, Application US/08477326
Patent No. 5968769
GENERAL INFORMATION:
APPLICANT: Green, Bruce A.
APPLICANT: Britton, Jr., Charles C.
TITLE OF INVENTION: Sequence and Analysis of LKP Pili
Patent No. 5968769

CORRESPONDENCE ADDRESS:
 ADDRESSEES: Hamilton, Brook, Smith & Reynolds, P.C.
 CHAIRMAN: M. J. Millett

:
 : COUNTRY: US
 : ZIP: 02173
 :
 : COMPUTER READABLE FORM:
 :

```

C:\>FC DOS/MS-DOS
C:\>
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/277,231
; FILING DATE: July 10 1994

REGISTRATION NUMBER: 33,542
REFERENCE/DOCKET NUMBER: ACC94-02A
TELECOMMUNICATION INFORMATION:

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

US-08-477-326-7

```

0Y      22 APVTKERALSLARLR-----QVSLMGOTAVSDKISIDYRROGDNERGITSOEEG 74
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

242 KVSFIVP-----FSNLAPLMRVGHURLYOLAGGRYDS-----PFDERVYLG 285

286 VLOYGLTNHLTNSLLYTRHYRAGLF-----GFGINTPIGAFSADATWSHAEFFPLKHVS 340

211 OEGGSEFYHAGSLVGNNSGTF--GEMYDWINNGPNDDLNLTSQYDAYATILMDYSASY 268

399 RPKKQFQVSLSSQSLGNNMGMLYLISGQTFYNWYMERGTN-----TQYGV-AVSNSFPHLNTSV 452

DB 453 NLSQSIDKRTGKRDNSIYLSLPLGDNHSDSSYSKSGNDINQRLGVNGSF--GERHGW 510
QY 315 INGEIASGRPVYAGNNGSIGHAFVCDGYAS-DGTFHFMNGMGVSGNGFKLLSTSL 373
DB 511 SYINMS-----RNNO-----GYRSDGNLSHNSISGY-RASYSRSLKNRSI 553
QY 374 GIGEGE-----GFTYOEIITGIPAKTPAAGTDALPILAKDIEAEKESGLNVG 427
DB 554 SLDGASGVAAHKIGITLSQV-----GESFALIHAKD-AAGAKYESGANNVS 598
QY 428 YSIVNTGEOSNLDLGYRLNKADGEVILEVKTSSINISWGY---GEHPESFSLAPNOLS 483
DB 599 LDYF-----GNAMVPTSPYEIYIGINPSDAEANEVEFATERQII 639
QY 484 OGINTITLXRTGTEOMEVPRAGGYSIVNTDPNNVY---TYDNNEGKLSIV 539
DB 640 PRANSTSLDFERGNK-----TMVLNLTLPNGEPVPASTADQDEG----- 681
540 PMSFVADL 547
682 --AFVGDV 687

RESULT 19
US-08-620-717A-9
; Sequence 9, Application US/08620717A
; Patent No. 5670365
; GENERAL INFORMATION:
; APPLICANT: Fellelson, Jerald S.
; TITLE OF INVENTION: Identification of, and Uses For, Nematocidal
; NUMBER OF INVENTION: Bacillus thuringiensis Genes, Toxins, and Isolates
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanichik & Saliwanichik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentlin
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/620,717A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/540,104
; FILING DATE: 06-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanichik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA94.C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (352) 375-8100
; TELEFAX: (352) 372-5800
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1168 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 167p
US-08-620-717A-9

Query Match 2.88; Score 125; DB 1; Length 1168;
Best Local Similarity 18.58; Pred. No. 0.5;
Matches 181; Conservative 147; Mismatches 368; Indels 282; Gaps 46;

QY 30 LSLARLARQVSLRMGOTAVSDKIS--IDVYRQGAERGITSOEESPAYVYVANNGN 87
DB 70 LALTKASISLILPGDAVAVPTIMFVDIFPR--LFRGSOQNAQAFELLIEKVK 126
QY 88 EGYALVAADR---IPITLAVSPIGRFDMSMDPLRMWLOIYDOEIGLLSKAQLNEE 144
DB 127 E--LVDEDEFRNTLNILNV-----LDGMOTALSHPN--DYQALIOGGEQPGMLMD 174
QY 145 ILRTEGVPAEVHALMDN-----GHPANDPMKNOGYPMNKEPL--LPNG 187
DB 175 QPTACTPTDHLISVRESFKDARTTIEFALPHKKNMLSTNDTPOFNSDTVLLTP-- 232
QY 188 NHAIVGCVATAAQAQIMKHYSPLOGESFPIYHAGSLVGNNSGTFGEYMDINMGNDLD 247
DB 233 --MT-----TAATLNLHOGYIO-----FAERKSVNY--DES 263
QY 248 NLTQSVDAVATATMRDVSASVMSFE-----NCSGTYSYVYGALRNPF-RYKRSIQH 301
DB 264 FINTKVDLQRR--QDYSTVSTTFEKKFPLNPSNKEV-----NKYNRYRSMTL- 314
QY 302 VRALYTSQEWHD-----INGELASGRPVY-----YAGNNSIGHAFVCDGYASDG 347
DB 315 --OSLDIAATPTLDNVPNSVNDIQDQTLVFSVAGPWEQNDNITSNI----- 363
QY 348 TFPHNMGMGVSGNGFYKLTLLSPSLIGGEGIGFTYOEIITGIPAKTPAAGTDALP 407
DB 364 -----IDVLTPLNTGIG-----FOE-----SSDKRTYTP 388
QY 408 ILALKDIE-----AEKSESGLNVGYSIYNTGEOSNLDLGYRLNKAD-GEVIEV 456
DB 389 RIELQSNQFQGYVNSKSVHEKCYSDGLKLVNKNKTITAGVNSIDESQNNKHNGPYINS 448
QY 457 KTSSINISWGY-----YEHPSFSLAPNOLSOGINTITLXRTGTE 499
DB 449 PTDINVNSQNSOYLDNSVMVNGGOKVACCSPLSSNGSNNAALPNQKINVIYSVQSD 508
QY 500 OMEPVRA-----OGYVNSIKVNTDPPNNVVYVDNNEKLSIYPSNFVD----- 546
DB 509 --KPEKHADYRKMGVYSSHLPYDLYPENYIGDIDPTKPSLLKFPRAEKGGSIDAY 566
QY 547 -----INSYESTITVOFNSDPEIRTPVAFALS-----TGATADD 583
DB 567 VSEPLNGANVKLTSYQ--VLKKEVNTQTKYRIRIRYATGDTASIMPHIIGPSGND 624
QY 584 VISLGYMAE-----VPGGSNTPYVMSKDVLTLSBGDTILMYTFSSINNQ-----KD 630
DB 625 LTNEGHNFSVSSRNKMFVQGNNGKYVNLITLDSIELPSCQOTILIO--NTNSODLFLDRI 683
QY 631 EMKKIGSVSKTPEYTHP-----LFEVGHNOSTYTLMAHNRVLPDFTLKNLGLPFN 684
DB 684 EFTSLPSTSTPTSTNFEPSLEKIINQVNOLESSTSSQOTELAH--TVSDKIIDOVVLKVN 741
QY 685 G-----ELVYVFRQTOSS-----SGSLWA-AQETVHIKGEFVYK 719
DB 742 ALSDDVGVGEKKALKRLVYNQAKOLSKARNVLVGSEFKGHEMLSRATVYVANNELKGD 801
QY 720 PVVEGP---IPDSYRATLHAFFVNGOOLYLKGRNTYK--IYNGTAVAEIES--SEI 772
DB 802 HLLPPTLTPSYAYOKI-----DESKLKSSTRYTVSGFIQSEHLEVVYSRVKEYV 853
QY 773 R-VFPNPARDYVEISAP-----CIPQETSIILFD--LSGKIYMNKSLASGHRDVS-RL 823
DB 854 HDMLDIPYEALPLISSDESPNCCPAACQCCSCGSDSHFYSIDVGSLOSDVNLGI 913
QY 824 PNGAVILKVDGYTTKINI 841
DB 914 EFGRIAKRNGFAKISNL 931

RESULT 20
US-08-296-791-3
; Sequence 3, Application US/08296791

```

Patent No. 6245337
GENERAL INFORMATION:
APPLICANT: St. Geme III, Joseph W.
APPLICANT: Falkow, Stanley
TITLE OF INVENTION: Haemophilus Adherence and Penetration
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Treccatlo, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1541 amino acids
TYPE: amino acid
TOPOLOGY: unknown
US-08-296-791-3

```

```

Query Match 2.88; Score 124; DB 4; Length 1541;
Best Local Similarity 20.48; Pred. No. 0.93;
Matches 197; Conservative 106; Mismatches 339; Indels 322; Gaps 53;

```

```

QY 57 YVYRQGDARIGTISOE-----GSPAYFYVANGNNE--GYALVAD 96
DB 210 FYKKGDNSLILNNEVGNNLKVGDATYTGIACTP---YKVNHNENGLIGFGNSKEE 266
QY 97 DRIPT-ILAYSPIGRFDM--DSMPDNLRMWLIQYDQIG--LILSG----- 137
DB 267 HSDPRGILSQDPLTNVAVLGDGSP-----LFVYDREKGMFLGSDYDFWAGYNKRSWQE 321
QY 138 ----KAOLNEELLRTGCVPAEVAEHALMDNGHFANDPMRKNQGYPW--NKEPILLPQGNHAY 191
DB 322 WHIYSQFTKVDVNLKDSAGSLIGSKTD-----YSWSNNGRTSTITGGEKSL 367
QY 192 TGCVATAAQAQIMRYHSMPLQEG-----SPDYHAGSLY-----GNWSG--- 229
DB 368 NYDLADGKDKRPHGKSVTFEGSGTLLNINNIDOGAGGLEFEGDYEVKGSNDTTPWKAGAV 427
QY 230 --TFEGMTDWMPCNPDLNLTQSQVDAYATL-----MRDVSASYSMSFYENG 276
DB 428 SVAECKTWTW--KVHNPQYDLAKI---GKGTLLIVEGTGDNKGSLLKVGDTYILKQTING 482
QY 277 SGTYSVVYVVAL--RNNFRYKRSQQLHVRALYTSQEHMDMIGELASGRPYTYAGNNOSI 334
DB 483 SGQHFASVSGTYSVSTVLNDKQVDPNSTYFG-----FRG---GR-LDLNNGNSLTF 531
QY 335 GHAFVCDGYSADGTGFHFMNGWGVSNFGYKLLLSPTSLGIGGEGIGFTYIOEITG--- 391
DB 532 DHIRNID----DGALVHNNMTNASN-----ITITGE-----SLITDPNT 567
QY 392 IEPAKTPAEAGTDALPILALDIEAEVKSGLNV--GYSIY-----NTGEE 436
DB 568 ITPYNIDAPDDBDNPYAFRRIRKDGOLY-----LNLENTYVALRKGASTRSELPRKSGES 622

```

```

QY 437 QSN-LDGLYLNKAD-----GEVIEYKSSINISWYGEHEPSFS 476
DB 623 NENMLYMGKTSDEAKRNVNHNINNERMNGFNGYFGEEGKNNGLNLTVEFKGSKQURFL- 681
QY 477 LAPNLSOGINTITLLYRRTGT--EQMEPYRHQGVYSIKVNTDP-----NNVYVYD 530
DB 682 ----LTGGINLNGDITVEKGTFLSGRPLPHARD--IAGISSTKQDPHAEENVEVED 734
QY 531 ----NNEKLSIVPNSFYADLNS-----YE-HSTITVO----- 558
DB 735 DWINRNFKATYTNVNTGNASLYSGRVANITSNTITASNKAQVHIGYGTGDTVCVRSYDTGY 794
QY 559 -----FNSDPDEIRTPVPAFAISTGATADVLSLGMVAEYVCGS-----S 599
DB 795 VICTTDLKSDKALNSFNPNLKNVNLTSANFVLKANKLFGTISR--GNSQVRLTENS 852
QY 600 NPPVYMSKVY--LTLSEGDYTLMYRPSINNOKDEMKGISVSQKPTPEX--THPLFEVGN 656
DB 853 HWHLTGNSDVHOLDLNGHHL-----NSADNSNNV-----TKYNTLVNSLSGN 897
QY 657 QSTYTLDAHNRVLPDFTLKNLGLPFNGELVYVFRQTOSSG--SLMAAQETVHIKQGE 714
DB 898 GSEYTLFDLSNKQ-----GDKTVV--TKSANGNFTLQVADTGPBNHNE 939
QY 715 TEVYK-----PYVEGPIDPGSYRATLHAFVNGQOOLYLK--GGRNTYKVIYNGT 761
DB 940 LTLFDASKRQRDHLNVSIVGNVYDVGAWYKILRN--VNGRYDLINPEVEKRNQTVDTNT 998
QY 762 AVEAIE-----SSEEL-RVFPNPARDYVEISAPCIPOETSILFDLS--GRIYVKN 809
DB 999 TPNNIQADVPSVSNNEETIARVDEAP---VPPAPAPPSETTEVEAENSQESKTEVKN 1054
QY 810 SLISA 813
DB 1055 EODA 1058

```

```

Search completed: May 29, 2003, 13:27:45
Job time : 23 secs

```

[illegible]

OY 219 ----HAGSLVGMGSGTGEYMDWIN-MPGNPDLDNLQSOVDATATLMRDVSAVSMSFY 273
 DB 223 PYFHPKRL---FAAISTROYNNMNLPTYSGRSNOVKM--AISELMADVIGSDVDYG 277
 OY 274 ENGSGTYSVYVVGALRNFRYKRSLOLHVRLALYTSQEMHDMINGELASGRPVYVYANNOS 333
 DB 278 PSSGSASSRVOALKEKNGYNSVHQINRGDFSKODMEAOIKELSONGPVYVYGK 337
 OY 334 IGHAFCVDCYASDGTFFHNMKGVSNGFYKLTLLSTSLGIGSGEIGFTTYOETITGIE 393
 DB 338 GGHAFVYIDGADGRNFTYHVMNMGVSDGFFRLDALNPSALGTGGAGGFGNGYOSAVYGIK 397
 OY 394 P 394
 DB 398 P 398

RESULT 2

US-10-002-784A-27
 Sequence 27, Application US/10002784A
 Application No. US20030036644A1
 GENERAL INFORMATION:

/33
 APPLICANT: Ulrich, Robert G.
 TITLE OF INVENTION: Bacterial Superantigen Vaccines
 FILE REFERENCE: 003/233/SAP
 CURRENT APPLICATION NUMBER: US/10/002,784A
 CURRENT FILING DATE: 2001-11-26
 PRIOR APPLICATION NUMBER: 08/882,431; 09/144,776
 PRIOR FILING DATE: 97-06-25; 98-09-01
 NUMBER OF SEQ ID NOS: 40
 SOFTWARE: Apple Macintosh Microsoft Word 6.0
 SEQ ID NO 27
 LENGTH: 468
 TYPE: PRT
 ORGANISM: Artificial sequence
 FEATURE:
 OTHER INFORMATION: mutant Spea-mutant Speb fusion
 US-10-002-784A-27

Query Match 8.9%; Score 393; DB 9; Length 468;
 Best Local Similarity 35.0%; Pred. No. 8.5e-22;
 Matches 106; Conservative 43; Mismatches 114; Indels 40; Gaps 13;

OY 111 FDMDSMPD-NLMMMLQIY-DOETGLISGAQLNEIETECVPAEVAHLMNDNGHFRANDP 168
 DB 184 FDFPEPEPTOSKYLMYKDN---TLDNTOI--EYLLTKOPV-VKSLD-----SKG 232
 169 MRMNOGYPMNNKEPLLPN-----GNHAYTGCVATAAQAQIMRYHSPLOGEGSFYD- 218
 DB 233 IYHNGNPNLTPYIEKKKPEOSFVGOHAATGCVATATAQIMRYHVPNNKGLDYTT 292
 OY 219 -----HAGSLVGMNMGSGTGEYMDWIN-MPGNPDLDNLQSOVDATATLMRDVSAVS 268
 DB 233 LSSNNPYFHPKRL---FAAISTROYNNMNLPTYSGRSNOVKM--AISELMADVIGSV 347
 OY 269 SMSFENGSGTYSVYVVGALRNFRYKRSLOLHVRLALYTSQEMHDMINGELASGRPVYA 328
 DB 348 DMDGPPSSGSASSRVOALKEKNGYNSVHQINRGDFSKODMEAOIKELSONGPVYVYQ 406
 OY 339 GNNOSIGHAFVDCYASDGTFFHNMKGVSNGFYKLTLLSTSLGIGSGEIGFTTYOET 388
 DB 407 GKGKYG-GHAFFVYIDGADGRNFTYHVMNMGVSDGFFRLDALNPSALGTGGAGGFGNGYOSA 465
 OY 389 ITG 391
 DB 466 VVG 468

RESULT 3
 US-10-002-784A-25
 Sequence 25, Application US/10002784A

Publication No. US20030036644A1
 GENERAL INFORMATION:
 /33
 APPLICANT: Ulrich, Robert G.
 TITLE OF INVENTION: Bacterial Superantigen Vaccines
 FILE REFERENCE: 003/233/SAP
 CURRENT APPLICATION NUMBER: US/10/002,784A
 CURRENT FILING DATE: 2001-11-26
 PRIOR APPLICATION NUMBER: 08/882,431; 09/144,776
 PRIOR FILING DATE: 97-06-25; 98-09-01
 NUMBER OF SEQ ID NOS: 40
 SOFTWARE: Apple Macintosh Microsoft Word 6.0
 SEQ ID NO 25
 LENGTH: 248
 TYPE: PRT
 ORGANISM: Artificial sequence
 FEATURE:
 OTHER INFORMATION: mutant streptococcal pyrogenic exotoxin-B
 US-10-002-784A-25

Query Match 8.7%; Score 386; DB 9; Length 248;
 Best Local Similarity 36.2%; Pred. No. 1.1e-21;
 Matches 93; Conservative 39; Mismatches 93; Indels 32; Gaps 8;

OY 155 VALMDNGHFRANDPMMNOGYPMNNKEPLLPN-----GNHAYTGCVATAAQAQIMRY 205
 DB 4 VASLID-----SKGHIYHNGNPNLTPYIEKKKPEOSFVGOHAATGCVATATAQIMRY 58
 OY 206 HSMPLGEGSFYD-----HAGSLVGMNMGSGTGEYMDWIN-MPGNPDLDNLQSOV 254
 DB 59 HNYPMKGLKDYTYLLSSNNPYFHPKRL---FAAISTROYNNMNLPTYSGRSNOVKM- 114
 OY 255 DAYATLMDVSAVSMSFYENGSGTYSVYVVGALRNFRYKRSLOLHVRLALYTSQEMHDM 314
 DB 115 -AISELMADVIGSVDMGPPSSGSASSRVOALKEKNGYNSVHQINRSDP-SODMEAO 172
 OY 315 INGELASGRPVYVYVYAGNOSIGHAFVDCYASDGTFFHNMKGVSNGFYKLTLLSTSLG 374
 DB 173 IKELSONGPVYVYGKYG-GHAFFVYIDGADGRNFTYHVMNMGVSDGFFRLDALNPSALG 231
 OY 375 IGEGSIGFTTYOETITG 391
 DB 232 TGGAGGFGNGYOSAVYG 248

RESULT 4

US-09-741-669-304
 Sequence 304, Application US/09741669
 Patent No. US20020022718A1
 GENERAL INFORMATION:
 APPLICANT: Forsyth, R. Allyn
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyskind, Judith W.
 TITLE OF INVENTION: Genes identified as required for
 FILE REFERENCE: ELITRA.009A
 CURRENT APPLICATION NUMBER: US/09/741,669
 CURRENT FILING DATE: 2000-12-19
 PRIOR APPLICATION NUMBER: US 60/173005
 PRIOR FILING DATE: 1999-12-23
 NUMBER OF SEQ ID NOS: 481
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 304
 LENGTH: 1325
 TYPE: PRT
 ORGANISM: Escherichia coli
 US-09-741-669-304

Query Match 3.5%; Score 154; DB 10; Length 1325;
 Best Local Similarity 21.7%; Pred. No. 0.0073;
 Matches 157; Conservative 86; Mismatches 260; Indels 220; Gaps 39;

OY 161 NGHFRANDPMMNOGYPMNNKEPLLPN-----PNCNHAFTGCVATAAQAQIMRYHSPLOGEGSFYD 217

Db 295 NGFNGCTVNISNNGLINNKESLVGVQDGSCHGVVNTDK----- 334
QY 218 YHAGSLVGNWS--CTGEE--MYDMINPBNPDLDNLQSOVDATATLMRDVSASVMSFY 273
Db 335 -----GHNFLGT--GEAFRIYIGDAGDELINVSSEKVD-----SGITIGMK 377
QY 274 ENGSGTYSV---YVVALRNNFRYKRSLOLVH--RALYTSQEMHDMIRGELASGRVYV 327
Db 378 ETGTGATTVKDKNSVITNLGTNLGYDGHGEMNINSGSLGCTGYETGVANVIT 437
QY 328 AGNNOSIG--HAFVCDGYASDGTFFHFMWG-----MGVSNGEFYKTLTTSPTS--- 372
Db 438 TGMAMEYKNVYTTIGVAGVGNLINSIDGKFVSQNTIFELGDKASGIGTLMNLMATSSFD 497
QY 373 -----LGIGEGCI-----GFTTYOE---ITIGIEPAKPPAEGTDALPTLAKIDEAEYS 420
Db 498 VGINVGFFGGIYVNSNGATLNSTGYGFIGSNAGKIVINISTDLNLTSSINACL-- 555
421 ESGLNMGYSIYNTGEOSNIDLGYRLNKADGEVEVETSSI-----NISWYGEHP 472
556 ---LQVG--VLGTGE-----LNTTGGIYKADTQIALNDKSKGVRDQNSL 600
QY 473 ESFSIAPNLSQGINITITLLYRRT---GTEQW---EPV-----RHAOGGYV 512
Db 601 ETENMYVG--TSGTGTLTNNGTLANVEGGEVYLGEFEPAGVTLNIGAHGEAADAAGFI 658
QY 513 -NSIKV-----NTDPN---NVVTVDNNEKLSIVPSPADLNSYHST 554
Db 659 TNAKTFEGLGEGVEFENHTNSDAGYQVMDLITGDDKDKVI-----HDA 704
QY 555 ITVQFNSDSPDEIRTPAFALSTGA--TADDVISLGWMAEV---PGSSNYPVWMSKD 608
Db 705 GHVFNAGNTYSKRLVNDGLLTASHTAGVGMG--SEVITIANG-----TLD 753
QY 609 VL--TLSEGDYTLMTRES---INQKDEMKKIGSVKPTTEYTHLFEVGHNOTSTYT 662
Db 754 ILASTNSAGDYTLTNALKGDLKGMVQLSSDKMEGFHATGEFA---GVAQLKDSFT 809
QY 663 LDMAHNRVLPDFLTKNLGLPFNGELVYVFPROTOSSSSLMAAGTVAHIKGEFFVY--- 718
Db 810 LEDDNRAALTHAMIQS-----DSENTTSVAVGEOSISGL-----AMNGTITIFPTDI 856
QY 719 --KPVVEGPIPDSDYRATLHAFVNGQOOLYLKGRNTYKIVNGTAVEALESSEIRVFP 776
Db 857 PAFTLAEGYI-----SVDTIYVAGADYTWKG--RANQ---VNGTGVLLD-----VP 898
777 NPARD 781
899 KPMND 903

RESULT 5
US-09-784-554B-2
; Sequence 2, Application US/09784554B
; Publication No. US20030032162A1
; GENERAL INFORMATION:
; APPLICANT: Schorr, Kirk
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schuelein, Martin
; TITLE OF INVENTION: FAMILY 44 XYLOGLOUCANASES
; FILE REFERENCE: 10017.200-US
; CURRENT APPLICATION NUMBER: US/09/784,554B
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1352
; TYPE: PRY
; ORGANISM: Paenibacillus polymyxa
US-09-784-554B-2

Query Match 3.3%; Score 148.5; DB 9; Length 1352;

Best Local Similarity 18.6%; Pred. No. 0.02;
Matches 176; Conservative 127; Mismatches 334; Indels 307; Gaps 47;

QY 112 DMDSPPNLKMLDIOYQETGL-----IISGAQLNEELRTEGV 151
Db 282 DMSVKNYSMFEDYDYLKQNSQVEGKRLLDVDFVHWYPEAMGGGIRITNEVGNDETK 341
QY 152 PAEVHALMDNGHFANDMFRNNGY-----PMNKK--EPLLP----- 185
Db 342 KARQA-----PRLMDPTYKEDSMIAQMSSELPILRLKQSVDRYTPCTKLAMT 392
QY 186 -----NGNHAATGCVATAAA-----OIMRHSWPLD--GEGSPDYHAGSLVGNWSG--T 230
Db 393 EYSYGENDISGGLAMTDVILGKNDVYMANMYKRLDGVANNVYSAAYKLYRNDGKNST 452
QY 231 FG-----EMTDNMGNPNLDNLQSOVDATATLMRDVSASVMSFYENGSGTY--SVY 283
Db 453 FGDTSVSAQTSDIYNSSVHASVTVNASDKELH--LVVNMKSMDSAPDAQFDLSGAKTYISG 511
QY 284 VVALRNNFRYKRSLOLVH--RALYTSQEMHDMIRGELASGRVYVAGNNOSI---GH 336
Db 512 VMGFDKNSQIKKAPITQISGRNFTTVP-----PLTAIYVLTGNDTSPVSGPE 563
QY 337 AFVCDGYASDGTFFHFMWG-----GWC-----GVSNG-- 361
Db 564 SFLKADAGOKVHLSDASGVYGVQVATDENGFTAVASNLFTSYDTNVTNGTS 623
QY 362 -FYKTLTSPSLGIGEGGFTTYOETITIGIEPAKPPAEGTDALPTLAKD--IARAY 418
Db 624 YYYKVT--ANTIKSSSSNLIKAVKMPVNG--PAREAEGET-----LKGTVSSG 672
QY 419 KSEGLNMGYSIYNTGEOS-----NLDLGR-----LN--KADGEVIE 455
Db 673 TGYSAGAYVNFHNPGLSLYTTIOAPTAGLNTLTIGRSPDOKRTNPSLNGKAFGELL 732
QY 456 VKTSINISWYGEHESEFSLAPNLSQGINITITLLYRRTGTBQWEPVRAAGGYNSI 515
Db 733 KKTADF-----KETSGKVLINAGANTISF--ETGM--GWYDI-----DYVRLE 772
QY 516 KVNTDNNVYVYDNNEKLSIVPNSFVAD-----LNSYEHSITVQFNSDSPDEIR 568
Db 773 PAADRPAAVATKLTN-----PNATVEKALMNTLVDDYGNMLSGOEIEIDMLO 824
QY 569 TPVAFALSTGATADVVISLGWMAE--VPGSSNYPVWMSKD--VLTLSBGDYTLWY--- 621
Db 825 ANV--GKPPAALDLIDYSPRAEHGLSTFEAKALAMQOGGIVFA-----WHMNA 876
QY 622 -RESINNO--KDEWKIGSVKPTTEYTHLFEVGHNOTSTYTLDMAHNRVLPDFLTKNL 679
Db 877 PKGLIDTQGEKEMWGFYADSTTFDIEYA---MNHPESEDKILLRIDIVIAG--QLKTL 930
QY 680 GLPFNGELVYVFPROTOSSSGS--LMAAOEYHIQO----- 712
Db 931 Q--DARVPVYLFRLHAEGRKFWMGAKGPEPVKLYILMHLRLTNVHKLNNLIWVNSV 987
QY 713 -----GEPFY-----YKPPV-----EGPIPD-----G 729
Db 988 APDWYRDEYVDILSPSYQAGDYSPQISKYEDELVALGDKRKLYVASENKPIDPDLMK 1047
QY 730 SYRATLHAFVNGQOOLYLKGRN--YVAKIVNGTAVEALESSEIRVFPNPARDYVETIS 786
Db 1048 AYQAHMWFATWYDGFVRDQKQNSLEHLKRYNHPNYITLDE-----LPTNLKTYGITE 1101
QY 787 APCIPQETSITLPDLGKIYKNSLSGHHMDSR--LPRGAY 828
Db 1102 QPSVPGSFTLMAAGETAKVSLSWTASANASYEVKRSRTTENGAF 1145

RESULT 6
US-08-837-459-19
; Sequence 19, Application US/08837459
; Patent No. US20020006407A1
; GENERAL INFORMATION:

```

APPLICANT: McKee, Marian L.
APPLICANT: O'Brien, Allison D.
APPLICANT: Machtel, Marian R.
TITLE OF INVENTION: Histidine-Tagged Intimin And Methods Of
TITLE OF INVENTION: Using Intimin To Stimulate An Immune Response And As An
TITLE OF INVENTION: Antigen Carrier With Targeting Capability.
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESS: Flanagan, Henderson, Farbow, Garrett &
ADDRESS: Dunner, L.L.P.
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/837,459
FILING DATE: 18-APR-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Forbids, Jean B.
REGISTRATION NUMBER: 32,984
REFERENCE/DOCKET NUMBER: 04995.0023-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 934 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-837-459-19

```

Query Match 3.2%; Score 140; DB 8; Length 934;
Best Local Similarity 17.4%; Pred. No. 0.05;

Matches 167; Conservative 128; Mismatches 331; Indels 336; Gaps 45;

```

74 GSPAYFYVANR--GNNEGVALAADRITITLAYSIGR-----DMSMPD-N 119
26 GIGLFFVYVONNSFANGENTFKLGSOSKLTTHDSYO--NLTFTLKTGTADLSKSDIN 83
120 L-RMW--LQIYDOEIGLILSGKAQNEIILRTGVPAPVHALMDNGHAMPKMGY 175
84 LSTIMELNKLHLYSSSEEMKKAAPGQ--QIILPLKLLPEYSAL-----124
176 PNNKNEPLLPNG-----NNAVTCGVATPAAQIMRHSMPLOG 212
125 PLLGSAPLVAAGVAGHTNKLTKMSPDYTKSNMTDKALNTYAAQOASISGLSRLNG 184
213 EGSFYHAGSLVGNNSGTFGEMYDWINMGPNPL-----DNLTQOVAVATLMDVSAS 267
185 DYAKDTALG-INGNASS--QIQANLQHYGTAEVMLGSGDNFGSSLD-----229
268 VSMSEYEN-----GSGT-----YSYVY-----284
230 FLPLPYDEKMLAFQGVAGARYIDSRFTANLCAQGRFLLPAMMLGVNFTDDFGSDNRL 289
285 -VGAIRNNFRYKRSQQLVHALYLSQEHMDIRGELASGRVY-----YAGNNSIG 335
290 GIGGYMDYFKSS-----VNGYFMRWRHESYHKRKYDERPANGFDIRFNGYLPSPALG 345
336 HAFVCDGASDGTFFHFMGMGVANGFYKLTLLSPTSIGIGEGIGFTIYOETITTEPA 395
346 AKLIEQYGGDNVALFNSD-----KL-----QSNPGAATVAVNTTPIPLVTMGI---389
396 KTPAEAGIDALPILALADIEAEVYSESGLANVYSIYVTGEOS--NLDLGYRLNKADGE 452

```

```

DB 390 -----DNR-----HGTGENDLISMGRFOEDFSWQ 417
OY 453 VIEWKTSINISMTGGEHEBSFSLAPNLSGIGINTITLLYR-----TGTEQ 500
DB 418 QIEPO-----YVNEURLTSGSRYPDYORNNHIIIEYKKODILSINIPHDINGTEH 467
OY 501 -----W-EPVRAAGGVYVSIIVNTTDPNNVYVTDNNEGKLSYV 540
DB 468 STQKIDLIYKSKYGLDRYVWDDLSALRSQGGQIDHSGSQAODYQATLPAYVGG--522
OY 541 NSFEADLNSTYHS-----TTVOFNSDSPDEIRTPVAFALSTGATADVISLGYW 591
DB 523 NIYKVTARAYDRNGNSNNVQOLTIVLSNGOVVDQVGTDFADTSAKADADITITN 582
OY 592 AEVPGG--SSNPVYVSKVVLISBEDYILMTFRFSINNOCKDKIGSVKTPPEYHP 649
DB 583 TYAKKGVAAQNVFVSFN-----IVSGTATL-----GANSKATDANGKATVTLKSTP-----629
OY 650 LEEVGHNOTSTYTLDMANHRVLPDFTLKNLGLPFNGELVVPVROTSSGSLMAAQETVH 709
DB 630 -----GQVVASAKTAEMSS-----ALNASAVIIPFDOKASITEL-KADKTTA 670
OY 710 IKOGE--TFVYKPPVEG-PIPDGSTR-ATLHAFVNGQOQLY-----K 748
DB 671 VANGKDAIKYTVKMKNGQPVNNOGVSTESTNGMFNGRSQCATGTGDRATITLTSSA 730
OY 749 GRNNTYKLVNCTAVEALESS--EIRV-----FPAPADYVISAFC--789
DB 731 GRATVSATVSDGAEVKATEVFPEDLKDKNVDIIGNVRELRINWLOTFQFLKASGG 790
OY 790 -----IQETSIIILFDLSGKIYKNSLSAGHRNDVSRPLNG-----AYILKYDGYTK 838
DB 791 DGTYSWYSEMTSIATVDASGKVTIN-----GKGSV-VIKATSGDKQVSYTIKAPSWIK 844
OY 839 IN 840
DB 845 VD 846

```

RESULT 7

US-10-080-505-11
Sequence 11, Application US/10080505
Publication No. US20030073166A1

GENERAL INFORMATION:
APPLICANT: St. Gene, Joseph W.

TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
FILE REFERENCE: A-59941-1/RT/DOE/DHR

CURRENT FILING DATE: 2002-02-22
PRIOR APPLICATION NUMBER: US 10/080,505

PRIOR FILING DATE: 1994-10-25
PRIOR APPLICATION NUMBER: US 08/296,791

PRIOR FILING DATE: 2001-04-20
NUMBER OF SEQ ID NOS: 58

SOFTWARE: PatentIn version 3.1
SEQ ID NO 11

LENGTH: 1391
TYPE: PRT

ORGANISM: Haemophilus influenzae
US-10-080-505-11

Query Match 3.1%; Score 138.5; DB 9; Length 1391;
Best Local Similarity 20.4%; Pred. No. 0.12;

Matches 165; Conservative 89; Mismatches 275; Indels 281; Gaps 44;

```

OY 67 GITSQEBGSPAYF-----YVANGNNEGVALAADD-----RIPTILAYSPIGRPD 112
DB 18 GIVSOAMAGHTYFGIDYQYRDFAEKRGKFTVGAQOIIDYKNGKEMIGTMKMGVMP--D 75
OY 113 MDSM-----PDNRKMLQI--YDOEIGLIL 135
DB 76 LSSMVGYSSTLISDEOHLISVAHNVGYDVVDFGMEGENDDHRRFYKYVKKRYNK-----130

```


Query 136 SGAQOLNE-----EILRTGVPAPAEVHALDNGHPAND-----PMRNOCYPM-----NNK 180
Db 131 SGRQYNDYOHPLERFETETABIEMVSYDGNHYKFNQYPLRVRGSHQWKKDDNNK 190
QY 181 EPLLPNGNHAATGCVATAAQMIRYHSMPLQGESEFVYAGSLVGNMSTGFEEMYMIM 240
Db 191 -----TIGDLAYGG-----SMLIGMTFEDGPGAGN----- 215
QY 241 PGAPDNLQSQVDATLMDRVASVSMSEFENGSGTYSYVVGALRNFRYKRSLOL 300
Db 216 -GTELENGRYQNP-NKYGPL-----PTAGSFDSGSPM-----FYDKYV-- 253
QY 301 HVALYTSQEMHDMIRGELASGRPVYAGNNSIGAFVCDGYAS--DGEFHFMWGVGS 359
Db 254 -----KKM--LLNGVLREGNRYAAGNSYQITRKDYFOGILNODITANF--WD--T 298
QY 360 NGFYKLLTSPSLGIGEGIGFTYQEI-ITGIEPAKTPAEAGTALPLALK--DIEA 416
Db 299 NAEYRF-----NIGSDHNGRVATIKSTLPKKAIOPERIYGLVDSQLHDAKNGDESP 352
QY 417 EYKSES-----GLNAGSYTYTGEQSNLDGYRLNKADGEV-----LEVTSINISWYG 467
Db 353 SYKGPWSPALHGHKSIYFGDGTGLTIENNINOGAGGLYEGNRYVVGNNQNTWOG 412
QY 468 YG-----EHPESFSLAP-----NOLSGINTITLLYRRGTG 499
Db 413 AGVSVGESEVEMOVHNPEDRLSKIGLGLTVNGKKNIGSLVNGVLV--DQADE 470
QY 500 QWEPVRAOGGYV--SIKVT--TDPNNVYV-----TVDNNEGLSTI--VPN--S 542
Db 471 SGQKQAFKEVGIYSRATVOLNSADQVDPNNIYFGRGRLDLNGHSLFERIQNTDEGA 530
QY 543 FVADLNSYESTITV-----QFNSDSPDEITPVAFALSTATADVLSLWV--MAEV 594
Db 531 MIYNHNASQJANTITIGMATINSKOL-----TNKKDIAFNCFEGODAKT 578
QY 595 PGG-SSNYPVWMSKDYLTLSGDTYTLWRESINNOKDEMKKIGSVS--KTPTETHP 649
Db 579 NGRLVNRYQPVNAENHLLSG-----TNLNGNITONGTGLVFSGRPPHAYNH- 627
QY 650 LFEVGHNGTSTYTLDMANRVLV-----DFTLKNLGLFPNGELVYVFRQTOSSS 698
Db 628 -----LRRDLSNMEGIFQGEIYWDHDMINRFEKAENFOIKGSAAVSRVSSIE 676
QY 699 GSILMAQETVHIKOGETFYVKKPYVEGPIPD 728
Db 677 GN-W-----TVSNNAATF-----GVVFN 694

DLT 8
US-10-080-505-15
; Sequence 15, Application US/10080505
; Publication No. US20030073166A1
; GENERAL INFORMATION:
; APPLICANT: St. Gene, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RT/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080, 505
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/286,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patent version 3.1
; SEQ ID NO 15
; LENGTH: 1391
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-080-505-15
Query Match 3.1%; Score 138.5; DB 9; Length 1391;

Best Local Similarity 20.4%; Pred. No. 0.12;
Matches 165; Conservative 89; Mismatches 275; Indels 281; Gaps 44;
Query 67 GITSQEBGSAFY-----YVANGNEGVALVAAD-----RIPTLAYSPIGRFD 112
Db 18 GIYSQAMAGHTYEGIDYQYRRDEAENKGFETGAODIDYKKGKEMGTMMKVPM--D 75
QY 113 MDGM-----PDLNMMWLOI-----YDQELGLL 135
Db 76 LSSWNGCYSTLISEOHLISVAHNGVDVDFGMEGNEPQHRFKYVRYNKK----- 130
QY 136 SGAQOLNE-----EILRTGVPAPAEVHALDNGHPAND-----PMRNOCYPM-----NNK 180
Db 131 SGRQYNDYOHPLERFETETABIEMVSYDGNHYKFNQYPLRVRGSHQWKKDDNNK 190
QY 181 EPLLPNGNHAATGCVATAAQMIRYHSMPLQGESEFVYAGSLVGNMSTGFEEMYMIM 240
Db 191 -----TIGDLAYGG-----SMLIGMTFEDGPGAGN----- 215
QY 241 PGAPDNLQSQVDATLMDRVASVSMSEFENGSGTYSYVVGALRNFRYKRSLOL 300
Db 216 -GTELENGRYQNP-NKYGPL-----PTAGSFDSGSPM-----FYDKYV-- 253
QY 301 HVALYTSQEMHDMIRGELASGRPVYAGNNSIGAFVCDGYAS--DGEFHFMWGVGS 359
Db 254 -----KKM--LLNGVLREGNRYAAGNSYQITRKDYFOGILNODITANF--WD--T 298
QY 360 NGFYKLLTSPSLGIGEGIGFTYQEI-ITGIEPAKTPAEAGTALPLALK--DIEA 416
Db 299 NAEYRF-----NIGSDHNGRVATIKSTLPKKAIOPERIYGLVDSQLHDAKNGDESP 352
QY 417 EYKSES-----GLNAGSYTYTGEQSNLDGYRLNKADGEV-----LEVTSINISWYG 467
Db 353 SYKGPWSPALHGHKSIYFGDGTGLTIENNINOGAGGLYEGNRYVVGNNQNTWOG 412
QY 468 YG-----EHPESFSLAP-----NOLSGINTITLLYRRGTG 499
Db 413 AGVSVGESEVEMOVHNPEDRLSKIGLGLTVNGKKNIGSLVNGVLV--DQADE 470
QY 500 QWEPVRAOGGYV--SIKVT--TDPNNVYV-----TVDNNEGLSTI--VPN--S 542
Db 471 SGQKQAFKEVGIYSRATVOLNSADQVDPNNIYFGRGRLDLNGHSLFERIQNTDEGA 530
QY 543 FVADLNSYESTITV-----QFNSDSPDEITPVAFALSTATADVLSLWV--MAEV 594
Db 531 MIYNHNASQJANTITIGMATINSKOL-----TNKKDIAFNCFEGODAKT 578
QY 595 PGG-SSNYPVWMSKDYLTLSGDTYTLWRESINNOKDEMKKIGSVS--KTPTETHP 649
Db 579 NGRLVNRYQPVNAENHLLSG-----TNLNGNITONGTGLVFSGRPPHAYNH- 627
QY 650 LFEVGHNGTSTYTLDMANRVLV-----DFTLKNLGLFPNGELVYVFRQTOSSS 698
Db 628 -----LRRDLSNMEGIFQGEIYWDHDMINRFEKAENFOIKGSAAVSRVSSIE 676
QY 699 GSILMAQETVHIKOGETFYVKKPYVEGPIPD 728
Db 677 GN-W-----TVSNNAATF-----GVVFN 694

RESULT 9
US-10-090-624-6
; Sequence 6, Application US/10090624
; Patent No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshi
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090, 624

CURRENT FILING DATE: 2002-03-06
 PRIOR APPLICATION NUMBER: 09/445,472
 PRIOR FILING DATE: 1999-12-06
 PRIOR APPLICATION NUMBER: 151969/1997
 PRIOR FILING DATE: 1997-06-10
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 6
 LENGTH: 1398
 TYPE: prt
 ORGANISM: Pyrococcus furiosus
 US-10-090-624-6

Query Match 3.1%; Score 138; DB 12; Length 1398;
 Best Local Similarity 19.5%; Pred. No. 0.13;

Matches 157; Conservative 91; Mismatches 262; Indels 296; Gaps 41;

185 PNGNHAYTGCVATAAQMIRHSHWPLQSGSPFDYHAGSLVGNWMSGTFGEYMDINMPGNP 244
 351 PNCGEYAVFG-----WDGHHGT--HVAQYAGYDSN--NDAMPMLSYSGE 392
 245 -----DIDNLTQ-----SOVDATATLMRDYAS----- 267

393 WEVFSRLXGMDYNTVTTDTVGAVAPCAQIMAIRLSDRGSMWDIEGMYAATGADY 452

268 VSMSP-----YENGSGTYSVY-----VG 286

453 ISMSLGNAPYLDGDPREVADELKEKGVYVIAAGNEPCINVSQVATKATITYG 512

287 ALRNNFRKRSIQLVHVALYTSQEHMDIRGELASGRPY-----AGNNOSIGHAFVC 340

513 A-----AAVPINV-GYVYSQ-----ALGPDYGFYFPATYRIAF-FSS 552

341 DGYASDGFHFHW-----GMGCVNGFYKLLLSPTSLSIGSGEGIFTYGEITIGEPAT 397

553 RGPRIDELKPNVAVAPG-----YSSLSLPMWIGG-----ADPMSGTSMAT 593

398 PAEACTDALPILALDIEAEYKS-----ESGLNVGYSIYNTGEEQSNLDLGYR----- 445

594 PHSVGVALLISGAAGIYVNPDIKKVLESGAWMBEDPYTGQKYTELDGQHGILVNT 653

446 -----LNKADGEVIEKTSINISWYGEHPESSSLAPNQLSOGINTITLLYRTGIE- 499

654 KSMELKAINGTTLPIVDHMAKSYSDAEV-----LGVYIRGIRYARNSIPD 701

500 --QMEPVRAAGGYNSIKVNTTDP-----NNVVYVDNNEGKLSIYNSFVADLSYE 551

702 IYEMH-IKVGDEYRTEIYATEPMIKRPFVSGVILENTEFLV-----KYDEGLE 755

552 HSTITVQFNSSDPDEIRTPVAPALSTGATADDVISLGWMAEVPGGSSNYPVWMSKDYLT 611

756 PGLYVGRILLIDP-----TTPV-----IEDEIL-----NTIYIPEK---F 787

612 LSEGYTL-WYRFSINNOKDEWKIGSVKPTTEYTHPLFEVGNOSTYTLDMAHNRV 670

788 TPENNYTTLWY-----DINGPEVTHHFTVPBGVLYLMTTYWYG 830

671 L--PD-----FTLKNL-----GLPENGELVVV--FROTQS-----SGSLW 702

831 LYRPGMVFYQDLYLPAAVSNRPMGMLWTFNFAPLYESGFLVRIYVELTTPSW 890

703 AAQET-----VHIKGETFYKPVVEG--PIPGSYRATLHAFAVNGOOLLYLKGRNY 753

891 YINRYLDTNTEFSIEFMITNIYAPINATLLPIGIGYNASVESV--GDGEFFIKG----- 944

754 TVKIVNGAVAEIIESESEIRVFPNARDYVELSACIF-QEISITLFLDLSGIYVKNKSL 812

945 -----IEVPSTAEK-----IRGNPSVNSDLDTLYLSKGLVALDGNP 986

813 AGHRMDVSRLLPNCAYILKVDGYTK 838

987 TAEEVVEYPRKPGYSIVHGYSVR 1012

RESULT 10
 US-10-080-505-17

Sequence 17, Application US/10080505
 Publication No. US20030073166A1

GENERAL INFORMATION:
 APPLICANT: St. Gene, Joseph W.

TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS

FILE REFERENCE: A-59941-1/RT/DCF/DHR

CURRENT APPLICATION NUMBER: US/10/080,505

CURRENT FILING DATE: 2002-02-22

PRIOR FILING DATE: 1994-10-25

PRIOR APPLICATION NUMBER: US 08/296,791

PRIOR FILING DATE: 2001-04-20

NUMBER OF SEQ ID NOS: 58

SOFTWARE: PatentIn version 3.1

SEQ ID NO 17

LENGTH: 1411

TYPE: prt

ORGANISM: Haemophilus influenzae

US-10-080-505-17

Query Match 3.1%; Score 138; DB 9; Length 1411;
 Best Local Similarity 18.8%; Pred. No. 0.13;

Matches 180; Conservative 124; Mismatches 325; Indels 330; Gaps 48;

1 MKK-----SFLAIYALFETIAMQ-----GHSAPYTKERALSIALRLRQVSLRNGOTAVSDK 52

1 MKKTYERLNFELACISL-GIVSQAMAGHT-----YFGIDYQYRDFENRGRKFTVGAK 52

53 ISIDVYRQGDAREGRTSQEESGSPAFYVANGNNGEYALVAADRIPIT--LWYSPIG 109

53 -NIEVYNNKGNLVG--TSMTKAPMIDFVSVR--NGVALVGDQYIVSAVAINGTAND 106

110 RFDMDSPDLRMLQI-----YDQELGLISGRAQNEILRTGEPAYH 156

107 FGAEQNPQOHFTFYKIVARNYKNDQTHPYEKDY-----HNPRLHKFV--TEAPRIDWT 159

157 ALMDNGHFAN-----DPMKMNQGYR-WNNKEPDLRPNHAYTGCVATAAQMIRHSHWPL 210

160 SDMNKNKTYDRKYDERNINISGQWQFWRNDQ-----NNGD----- 194

211 QEGSGFDY-----HAGSLVGNMSGTFGEYMDINMPNPDLNLTOSQVDAVATLMRD 263

195 QVAGAYHYLLAGNTINOGAGGWSLSGD-----YRQGN-----YGPi--- 234

264 VSASVMSFYENGSGTYSYVVGALRNFRYKRSIQLVHVALYTSQEHMDIRGELASGR 323

235 --PIAGSSGDSGSEMF-----YDAEKOKWMLINGVLRIGN 267

324 PYVYAGNN--OSIGHAFQDGYASDGFHFEMWGMGVSNGFYKLLLSPTSLSIGSGE 380

268 P--WAGTENTFOLVAKSFDEILEKOLRTSF--YSPSGNGAYTTT-----DKGDS 314

381 GFTIYQ-----EITIGIEPAKTPAEAGTDALPILAKDIEAYKSES-----GLNVGYSIY 431

315 GIVKQQTGRPSFVRIGLMDKLPAG-----KDDVYQYQPNITLPLNNGNLY 364

422 NTGEQSNLDLGRANKADG-----EYIEKTSINISWYGEHPESSSLAPNQLSOGIN 487

365 FGDQNGYVTLSTINOGAGGLYFEGNFTVSSENNATWOGACVHGEDS----- 413

488 TITLLYRRTGTQ-----WEPVRAAGGYNSIKVNTTDPNNVYVYVDNNEG-----K 535

414 --TYWKYNGVENDRLSKIGKTLHVAKGKNGKISVGDGKVLLEQADQGNKQANSE 471

536 LSIYNSFVADLNSYESTITVQFNSD-----SPDEIRPV 571

472 IGLVSGRGTYQLNDK-----QFNTDKFYFGRGRGLDNLGHSILFKRIQNTDGAITY 525

572 APALSTGATADDVISLGWMAEVPGGSSNYPVWMSKDYLTLSGDTYTLMTRESIN--NQ 628

Db 526 NNNATSTV-----TITGSDTINDNTGDLTKNRDIAFNMGFGD 564
Qy 629 KQMKKIC--SVSVKPTETHPLEFVGNQSTYLLD-----MAHNRVLPDF 674
Db 565 KQDTKMTGRLMYNPLNDNHLGSGNLTGQDGTLYVSGRPYHAYNLH---N 621
Qy 675 TLKNIILPENGELV-----VFROTSSSGSLMAOETVHIKQOETFYKRVV--VEGR 725
Db 622 RLNELGRP-KGEVYIDDDMINRFX-----AENFOIKGSGTYSRVNVSIEGN 668
Qy 726 -IPDGSYRATLHAFVNGQOQLYLGKRN--YTVKIVNGTAVEALEESSEIRVFPNPARDY 782
Db 669 WTISNNANATFGVVPNOQNTICRSDMTGLTCTKTVNLTKDKVIDS-----714
Qy 783 VESAPCIPOETSIILFDLSGKIYMKNSLSAG-HGRMDYSRLPNNGVILKVDGTYTKIN 840
Db 715 -----IPTT-----QINGSTINLTNNATVNIHG-----LAKNGNVTLIN 748
Query Match 3.1%; Score 136.5; DB 9; Length 3354;
Best Local Similarity 20.1%; Pred. No. 0.64;
Matches 199; Conservative 126; Mismatches 338; Indels 327; Gaps 52;
Sequence 12, Application US/10160758
Publication No. US20030036076A1
GENERAL INFORMATION:
APPLICANT: EXELIXIS, INC.
TITLE OF INVENTION: CADS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE
FILE REFERENCE: EX02-089C
CURRENT APPLICATION NUMBER: US/10/160,758
CURRENT FILING DATE: 2002-06-03
PRIOR APPLICATION NUMBER: US 60/296,076
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/328,605
PRIOR FILING DATE: 2001-10-10
PRIOR APPLICATION NUMBER: US 60/357,253
PRIOR FILING DATE: 2002-02-15
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Patentin version 3.1
SEQ ID NO 11
LENGTH: 3354
TYPE: PRT
ORGANISM: Homo sapiens
US-10-160-758-11

Query Match 3.1%; Score 136.5; DB 9; Length 3354;
Best Local Similarity 20.1%; Pred. No. 0.64;
Matches 199; Conservative 126; Mismatches 338; Indels 327; Gaps 52;
Sequence 12, Application US/10160758
Publication No. US20030036076A1
GENERAL INFORMATION:
APPLICANT: EXELIXIS, INC.
TITLE OF INVENTION: CADS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE
FILE REFERENCE: EX02-089C
CURRENT APPLICATION NUMBER: US/10/160,758
CURRENT FILING DATE: 2002-06-03
PRIOR APPLICATION NUMBER: US 60/296,076
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/328,605
PRIOR FILING DATE: 2001-10-10
PRIOR APPLICATION NUMBER: US 60/357,253
PRIOR FILING DATE: 2002-02-15
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Patentin version 3.1
SEQ ID NO 12
LENGTH: 3354
TYPE: PRT
ORGANISM: Homo sapiens
US-10-160-758-12

Db 631 -----QISNGILYLVFM---AMDAGNPPLNSTVPTVTEFDENDNPTEFSKP 674
Qy 400 -----EACTDLPILALK-DIEAEKSSGLNVCSTYNTGEBOSNLDGRLNK 448
Db 675 AYFVSVENIMAGATVLFNATDLDLRSREGQBS---ITSL--EESTO-----PRIN 723
Qy 449 ADEEVI-----EVTSSINISMY---GYGEHPRESPLAPNOLSGINT--ITLLYRRT 496
Db 724 RSEIITTSILDBETSEVILYIYRAVDGCGHOKT-----GIATVNTLLDIND 773
Qy 497 GTBQWEPVRAHOGGYVNSIKVNTDPN-NVYVTVDNNEK-----LSIYPSFVADLNSY 550
Db 774 NHPYTDAPY-----YINLVEMTPPDSQVTVVAVDPLGNGTLVYSIQPNKFYSINST 829
Qy 551 --EHSITVOFNSDSD-----EIRPVAALSTGATADVLSL-----587
Db 830 TGRIRTHAMLDRENDDPHEALMRRIYVSVCGRPLKATSSATVFNVLDLNDNPT 889
Qy 588 -----GWMAFVPGSSNYPVWMSKDVLTLESDYTLWYRPSINNOKDEMKGISVS 640
Db 890 FQNLPRVAEVLBEIPAGVSTYQV-----AIDDEG-----GLVSY 927
Qy 641 KPTETHPLEFVGNQSTYLLDMAHNRVLPDFTK---NLGLPENGELVYVFKQTS 696
Db 928 RMPVGMPRMDFLINSSGVVVTTELDREIAEQVLRVVASDAGTP-----TKS 976
Qy 697 SSGSLMAOETVHI--KQETF---YKPYVEGRIP-----DGSYRATLH 736
Db 977 STSTL-----THVLVDNDETPFPFAVYVNSVEDPREFRVWMLNCTDNDGLNAELS 1031
Qy 737 AFVNG--QOOLYLGKRNKYVIVNG-----TA-----VEAISSEIRVFPNPARDYV- 783
Db 1032 YFTTGGVNDGKESVGRDAVRYVGLDRTTAYMYLLEAIDNGPRKGRHGTATYFV 1091
Qy 784 -----EISAP-CIPOETSIILFDLS---GKI--YMKNSLSAGHG---R 817
Db 1092 VLDVNDNRPFLQSSYEASVPEDIPGHSILQKATADGEGRGVYRILHGNHGNFR 1151
Qy 818 MDYS-----RLPNGAVILKVDGY 835
Db 1152 IHVSNGLMGRPPRLDRERNSSHVLYEAY 1181

Query Match 3.1%; Score 136.5; DB 9; Length 3354;
Best Local Similarity 20.1%; Pred. No. 0.64;
Matches 199; Conservative 126; Mismatches 338; Indels 327; Gaps 52;
Sequence 12, Application US/10160758
Publication No. US20030036076A1
GENERAL INFORMATION:
APPLICANT: EXELIXIS, INC.
TITLE OF INVENTION: CADS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE
FILE REFERENCE: EX02-089C
CURRENT APPLICATION NUMBER: US/10/160,758
CURRENT FILING DATE: 2002-06-03
PRIOR APPLICATION NUMBER: US 60/296,076
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/328,605
PRIOR FILING DATE: 2001-10-10
PRIOR APPLICATION NUMBER: US 60/357,253
PRIOR FILING DATE: 2002-02-15
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Patentin version 3.1
SEQ ID NO 12
LENGTH: 3354
TYPE: PRT
ORGANISM: Homo sapiens
US-10-160-758-12

344 NAEFNSSESVATITELAGVAFALPLEIYVVK-----DENGLSNFEVLYGN 393
QY 76 PAFVYANRGNNGVYALVADDRIPTILASPIGRDMD-----SMDNI---RMLQIT 127
DB 394 NSHFTISPTSVGKADIRIARIP--LDEYDRYDFDLFANESVDPDHGAKVITLI 451
QY 128 DOEIGLILSGKAOLNEEILTEGVPAEVHALM--DNG-----HFANDPRWNOGY 176
DB 452 NENDNRPISQPLYNISLVENVGTSVLYLATDADAGFGEVYFSDPDRFSL--- 508
QY 177 MNKPELLRNGNNAIYGCATTA---AQMRYHSPLOGEGSFYHAGSLVGNMGTG 232
DB 509 --DKD-----TGLIMLARLDYELORFTLILARDGGEGTGRVIRIN----- 550
QY 233 EMTDWINMPGNPDLMLTOSQVAYATLMDVSAVSM-----STENSGTYSY 283
DB 551 -VLD-VN-----DVPFFQKADYAGALRENEPSTOLVRLATEDSDPPNQITYSIV 601
QY 284 VVGALRNFRYKRSLOLHVRAVLTSGEMHDMINGELASGRPVYAGNNOIGHAFYCDGY 343
DB 602 SASAFSGSYFDI-----SLEYG-----GVISVRPLDYE----- 630
QY 344 ASDGTFHFMGMGVNGFYKLLLSPTSLGIGEGITTYOELITIGEPKATPA----- 399
DB 631 -----QJNGLIYLVW---AMDAGNPLNSTVPVITIEVDENDNPPFSKP 674
QY 400 -----EAGTDLPIILAK-DIEAEYKESGILNVSITYNTGEOSNLDLGRANK 448
DB 675 AVFVSVENIMAGATVLFINTDLDLRSRYGOES--IYSL--ESTO-----PRINA 723
QY 449 AGCEVY-----EVTSSINISW---GYGHPESFSLAPNOLSGINT--ITLLYRRT 496
DB 724 RSGELITTSLLDREKSEYILIVRAVGGVGNOKT-----GIATVMTLLDIND 773
QY 497 GTEOMEVYHAGGYNSIKVNTDPN--NVVVTVDNECK-----LSIYPSFVADLNSY 550
DB 774 NHPTKADRY---YINLVEMPRPSDVTYVAVDPOLENGTLYISIDPPMKFYSLNST 829
QY 551 --EHSTTYQFNSDSD-----EIRTPAFALSTGATADVLSL----- 587
DB 830 TCKIRTHAMLDRENDDPHEALMRKIYVSTDGCRPLKATSSATVFNLLDINDNPT 889
QY 588 -----GVMVAEYVGGSSNPVWMSKVDLTLSBGDTLMTFRSTINQDEKKIGSV 640
DB 890 FQNLFPVAVELGCTPAGVSTIOV---AIDDEG-----LN-----GLVSY 927
QY 641 KTPLEYTHLPFVGNHNOTSTYTLDMAHNRLPDLTK-----NLGLPENGELVVFROTS 696
DB 928 RMPVGMPRADFLINSSGVVTTTELDREIRIAEQLRVVASDAGTP-----TKS 976
QY 657 SSGSILMAOEVYH---KQGETF---YKKPVYEGPIP-----DGSYRATLH 736
DB 977 STSTL-----TIHVIDVNDERTFPFPAVYNSVSDVREFRVVWLNCTDNDVGNALSL 1031
QY 737 AFVNG--OOQLYLKRNVTYKING-----TA-----VEATESSEELRVFNPARDV- 783
DB 1032 YFITGNNDGKRSVGRVAVVTVGLDBRETTAAMLIBEALDNDGPGVRHTGTAIVPT 1091
QY 784 -----EISAP-CIPQETSILFDLS-----GRI--VMKNSLSAGHG---R 817
DB 1092 VLDVNDNRPIFLOSSYEASVPEIDPEGHSILOLAKTADDEGEFGVWVRILHGNNGNFR 1151
QY 818 MDVS-----RLPNKAYILKVDGY 835
DB 1152 IHVSNGLMRGPRPLDRERNSSHVLIVEAY 1181

RESULT 13

US-09-738-626-6239
; Sequence 6239, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI

APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SETKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIRO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, MOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OKAI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 6239
LENGTH: 2993
TYPE: PRN
ORGANISM: Corynebacterium glutamicum
US-09-738-626-6239

Query Match 3.0%; Score 131; DB 9; Length 2993;
Best Local Similarity 19.5%; Pred. No. 1.4;
Matches 150; Conservative 87; Mismatches 258; Indels 274; Gaps 37;

QY 136 SGAOLNEEILTEGVPAEVHALMDNGHFRANDPMHNGGYPMNNKEPLLNPNNHA--YTG 193
DB 651 SGRSQLGADIHEDINDSFAKAGRLD--EVAGD-----EFAVQARDEITLAIKTAKYFG 704
QY 194 CVATAAOLMRYHSM-----PLGEGSPDYHAGSLVGNMGTGEMVD----- 236
DB 705 DIGS-----MTEYQWLMRYLELSGVPDQW-----IDASWAARFQMLERARALIE 751
QY 237 -----WJ-----NMPGNP-----DL 246
DB 752 ODHGPEPSLVEDVDKLVAAYPHAARDLLTPADVAFGLCRTGKRPVNEPVYIDRY 811
QY 247 -----DMLTOSQVAYATILMDVSAVSMSEYENGSGTYSYVVGALRNFRYKRSLOL 300
DB 812 RRMWRSDSLMOSSHDRYADQVAILPGV-----VAAGITKANDEPVADLLDR 858
QY 301 HYRALITYOEHMDM-----IRGELASGRPVYIAGNN-----OSIGHAFVCDGYASD--GTF 349
DB 859 FVDATIERIDEHDSRSDIMGKVLSSPGTFMGRNIPSVIHSLGHA---DWKSRSEFEAF 915
QY 350 HFNMG-----MGVSNGFYKLLTSLPTSIGIGE-GIGFTYOEITITIGEPKATPAEAGTD 404
DB 916 HSPICANLYEAEAEAMLTPLAGSTARGTTAEKIRFT-----SPIDALPS 962
QY 405 ALPILALDIEAEYKESGILNVSIV---YNTGEOSNLDLGRANKRADEVIEVKTSS 460
DB 963 AVPLVTOQDAEAAAMELFRIAAGTLATVNGTATWETSVAGY----- 1006
QY 461 INISWYGEHPRESLAPNOLSGOINTITLLYRTGTG-----OMEVYRHAOGGYVNSI 515
DB 1007 --IADY-----NNVTAGYLPASVYPAHTADVLVGRAMPVAFPA---VNSA 1047
QY 516 KVNITDPNNVYVTDNNEGKLSIYVNSVADLNSYEHSTIYVQFNSDPDEIRTPVAFAL 575
DB 1048 VITPDSASVY---EGMLSLV-----HEHHIYV---LKSVDVPTD-----GAL 1083
QY 576 STGATADDVIS---LGWVAEYVGGSSNPVWMSKVDLTLSBGDTLMTFRFSINQK- 629
DB 1084 KVSATADEVDVTDIGRLIYVAEITADAGNL-----IATLAE-----RFAIRGRKG 1129
QY 630 -----DEKKIGSVSVKTPITTYTHPLPEVGNHNOTSTYTLDMAHNRV-LPD 673

QY 213 EGSEFDYHAGSLVGNMSG---TFGEYMDWIMNPGNPDLNLTQSOVDAYAT----- 259
 DB 432 ANNVSAAYKLYLNDOCKSSTFGD-----ISVHAQTSIDLNVSSVHASTDASYKELHLY 485
 QY 260 -LMRDVSASYSMSFEENGSGTY-SVYVYCALRNFRKRSLOLHV-----RALYT---SOE 310
 DB 486 VMKSMDSADAOFDLSGETTYSGYKMGWEDKNSOIKKAAPIITQISGNFFYTVPLPLTA 545
 QY 311 WHDM-----IRGELASGRPVYVYAGNNOISIGHAFVCDGVSDDCTF----- 349
 DB 546 YHIVLVLGNPTVPENPESFALRAEADGKSIYLDASSGVYGVORATY-ENGPEFAVAS 604
 QY 350 ---HEPMWGVANG---FYKTLPLSPSLIGEGEIGFTIYOEITIGTEPAKTPAEAGT 403
 DB 605 NLVETSYTDNTNNGTSYKKT--AKTKGTASAVLKAVPAPVPG--PDVEAEDGT 660
 QY 404 DALPILALDIEAEYSESLANGY---SIYNGEEOO-----NIDLGR----- 445
 DB 661 ---LGTVESSGTFSGTGYVTFNHMGDSLMTIQAFTAGLYNLITGYSRPHD 713
 QY 446 ---LN-KADGEVIEVKTSSINISWYGEHRESEFSLAPNLSOGINTITLLYRTGT 498
 DB 714 KTNFSLNGKASELYLMKTADR-----KTSCKVLLNCAITIGF---ETG- 758
 QY 499 EOMEVPRHAOGY-VNSIKVNTT--DPNNVYTVNNEGKLSIYVNSFVAD-----L 547
 DB 759 --W-----GWTIDIDYVLEPADRPPHAYVTKTLIN-----PNATVEAKALNLY 801
 QY 548 NSEHSTITVOFNSDPDEIRTPVAFALSGATADVYISLGVMAVPPGSSVYVYVSK 607
 DB 802 DOYGRKMLSGO-----EMPEIDWLQANV-----GKPAIAL 834
 QY 608 DVLTLSEGDYTLMYRESINNOCKEKKISVSYPKTEYTHP--LEFEVG-----HNO 657
 DB 835 DLIDYSPRAEHLSTETEKALIMDKOGGI-VTFAMHNAKPGGLIDTGCKEMWGFYAD 893
 QY 658 TSTYTLDMAHNRYLPFTLKNGLPRNGELVYVFGTOSSSGSLMAAOT-----VH 709
 DB 894 STTFDELYAMNH-----PSESDYKLLIRIDIVLAGOLKLODAKVPVLEPRPH 941
 QY 710 IKOGEFVYKPVVEGPIPDGSYRATLHAFVNGOOLYLKGRNRYKIVNGTAVAEIES 769
 DB 942 EAEGRKFTWGG--AKGEPRYKLYILMH-----DLITNVHKLNLWVNSVAPDMVPGD 993
 QY 770 EDIRV-----FPNAPADYVEISAPCTIOETSIIPLDLSKIV 806
 DB 994 EYVDILSFDSPQ-AGDY-----SPQAKYEDLVLTGKDKKLY 1030
 DB 09-839-996-3
 Sequence 3, Application US/09839996
 Publication No. US20030009010A1
 GENERAL INFORMATION:
 APPLICANT: St. Geme III, Joseph W.
 TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
 STREET: 4 Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: United States
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/839,996

FILING DATE: 20-Apr-2001
 CLASSIFICATION: <unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/296,791
 FILING DATE: 25-AUG-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Treacartin, Richard P.
 REGISTRATION NUMBER: 31,801
 REFERENCE/DOCKET NUMBER: A-59941/RT/RMS
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELE: 910 277299
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1541 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 SEQUENCE DESCRIPTION: SEQ ID NO: 3:
 US-09-839-996-3
 Query Match 2.8%; Score 124; DB 9; Length 1541;
 Best Local Similarity 20.4%; Pred. No. 1.8;
 Matches 197; Conservative 106; Mismatches 339; Indels 322; Gaps 53;
 QY 57 YYRGGDAERGITSQEE-----GSPATYEVANRGNNE--GYALVAD 96
 DB 210 FYKKGKDNYSLLNHNHGGNNLKLVDATYTGIAETP--YKVHNENGLIGFENSKKE 266
 QY 97 DRIFT-ILAYSPIGRDM--DSMPDNLKMLQIYDOEIG--LILSG----- 137
 DB 267 HSDPGKILSDPLLYAVLVDGSGSP--LEVYREKGMWFLGSYDFWAGYNNKSMOE 321
 QY 138 ---KAQNEELLRTBEGVAEYAHALMDNGHFANDPRKNNOGYPW--NKEPLLPNGHAY 191
 DB 322 KNIYSSQFTKVDLNDSDAGSLIGSKTD-----YSWSNKSTSTTIGGEKSL 367
 QY 192 TGCVATAAQAQIRYHSMPLQEGE-----SPDYHAGSLY-----GNSSG--- 229
 DB 368 NVDLADGDKDPHNGSVTFEGSGITLNNNDGAGGLFEESDYVYKSTDNTHKKGAV 427
 QY 230 --TFGEYMDWIMNPGNPDLNLTQSOVDAYATL-----MRDVSASYSMSFEENG 276
 DB 428 SYAEGKTYTW--KVHNPQYDLRAKI---GKTLIVEGTGDNKSLKVDGTYILKQING 482
 QY 277 SGTYSVYVYVGA--NNNRKRSLOLHRYALYTSEHMDIRGELASGRPYIYAGNNOI 334
 DB 483 SGQHAFAVSGIVSGRSTVLANDKQVDPNSIYFG-----FRG---GR-LDLNNGNSLTF 531
 QY 335 GHAFVCDGYASDGTGTHFMWGVANGVNGFYKTLPLSPSLIGIGEGEIGFTIYOEITG--- 391
 DB 532 DHIRRID---DGALVHNHNTASN-----ITITGE-----SLTIDPT 567
 QY 392 IEPATPAEAGTALPILALDIEAEYKSEGLNV-GYSIY-----NNGEE 436
 DB 568 ITPYNIDAPDEDNPAFRIRKDGGLY---LNLNENTYVALRKAGSTRSELPRNSES 622
 QY 437 QSN-LDLGRLKAD-----GEVLEVTSSINISWYGYGERPESFS 476
 DB 623 NENMLYMKTSDEAKRNNYHNIINERANGFNGYGEEEKNGNGLNVTFKGKSEONREL- 681
 QY 477 LAPNLSOGINTITLLYRTGT--EOMEVPRHAOGVYNSIKVNTDP-----NNVYVYVD 530
 DB 682 ---LTCGTILNGLDLYEKTLPLLSGPTPHARD--IAGISSTRKKDPHFAENNEVVED 734
 QY 531 ---NNEGKLSIYVNSFVADLNS-----YE-HSTITVO----- 558
 DB 735 DWINRNEKATYMTYGNASLYSGRVANITSNTASNAQVHIQYKIGDTCVASHDITGY 794
 QY 559 ---FNSDSDERTRTYVAPALSGATADVYISLGVMAEYVPGS-----S 599
 DB 795 VTCYTDKLSKRALNSFNPTNLRGVYNLTESANFVLGRANLEGTIOSR--GNSOVRLTENS 852

QY 259 TLMRDVSAVSMSSEYENGSGTYSYVVGAL--RNNFRYKRSIQLHVALYTSOEHMDIRG 317
 Db 72 IWSSKXYIAAVEKAGVYVGDGKENEYFEGKIDIRASF-----ASMLVSAYNLKDKYVNG 122
 QY 318 ELAS-----GRPVYAGN---NOSTGHA---FV-----339
 Db 123 ELVTFEPDLDLHNGEERANILINLGISVGTGKWEKPKSVSRAMAFIALTRKRYKRD 182
 QY 340 -CDGASDGTFFHFMGWSNGFYKLLTSLGSGEGIGFTYIQTITGIEPAKTP 398
 Db 183 NAQAYVTD-----VKSEPTKLTITGDLKLSADDT--LEGKAY 222
 QY 399 A-EAGTALPLAL-----KDIEAEVSESGL--NVGSIYNTGEGOSNL-----441
 Db 223 AIBASDGTSAVYTLGKVPANRDLTVKKNOSFVTFEYKVKLAVERKLTFFDDDRAGQA 282
 QY 442 LGYRLNKADGEVIEVKSSINISWYGEHPSEFSLAPNOLSGINTITLLYRTGECW 501
 Db 283 IAKLNDEK-----NADVEYLANHDKF--VANNLDGSPANIEFGGERTSTTGK 332
 QY 502 EPVRAHAGGVNSIKV---NTDPNNVVYVDNNEKLSIVPSEFADLNSEHSTIYV 557
 Db 333 LAVGIRKGDYKVEQYTKRGGLVSNMTGITTAKNLDTPASAIKNVFA--LDADNDGVNY 391
 QY 558 QFNSDSDPEIRTVAPAL--STGATADDVSLGVMVAEVP-----GSSNTPVY 604
 Db 392 GSKLSGKD-----FALNSQNLVGEKASLKNLVATIGEDKRVDPGSISIKSNHGI 444
 QY 605 -----MSKDVLTISEGDTLWYRFSINNQDEMKISGVSKPTETHTHLEFV 653
 Db 445 SVVNNYITAEAGALITIVGDTYKDKFKVTDS---RKL--VSXKANPD---KLOY 495
 QY 654 GHNOSTYTL-----DMAHNRVLPDTLKNLGLPENGELVYVFRQTOSSSGS 700
 Db 496 VQNKTLPIVFTVTDQYGDPEGANTAIKELVP-----KTGVVAEGGLDV---TDSGS 546
 QY 701 LW-----AAQETVHKQGE---TFVYKPVYEGPDPGSIATATLHAFYNGQ---743
 Db 547 IGTGKTGVGNDVGEVTHQNGNATLSLYNVTEGVNAFKNFELSVKVGQSGSPPT 606
 QY 744 QLYLCKRMYTVKIVGTAVEALIESESEIRVFNPARDVYEISAPCIPOETSIILDLGS 803
 Db 607 KIDLAVSTVEYQLSTYSDRYVSDPENLEGYEVESKNLAVADAKIVGNKV--VYTKTIG 665
 QY 804 KI---VAKNSLSAGHRMDV 820
 Db 666 KVDIHLTKGATAGKATVEI 685
 LT 19
 Db 9-839-894-6
 Patent No. US2002017686A1
 GENERAL INFORMATION:
 APPLICANT: Alboum, Zeev
 APPLICANT: Barry, Eileen M.
 APPLICANT: Levine, Myron M.
 APPLICANT: University of Maryland
 TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
 FILE REFERENCE: UOFMD.006A
 CURRENT APPLICATION NUMBER: US/09/839,894
 CURRENT FILING DATE: 2001-04-20
 PRIOR APPLICATION NUMBER: 60/198,626
 NUMBER OF SEQ ID NOS: 40
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 6
 LENGTH: 867
 TYPE: PRT
 ORGANISM: E. COLI
 US-09-839-894-6

Query Match 2.8%; Score 122.5; DB 9; Length 867;
 Best Local Similarity 19.1%; Pred. No. 0.97;
 Matches 149; Conservative 105; Mismatches 238; Indels 287; Gaps 39;
 QY 43 RMGOTAVSDKISIDYVY-----ROGAERGITTSQEE--GSPAFVYVNRG-----85
 Db 230 RMDRTDLSGSGISGNFNENLLPLPDIDGIRGTQSYIKNDKRIAPVYVMLNFRIVERA 289
 QY 86 --NNEGVALYAADRIPTI--LWSPIGRDMSPONLKMWLOIYQOIGLLISGRAQIN 142
 Db 290 FRNNOLLGVWYLDGSGVNEIDTARLPYGSYDLK-----LKIFE-----N 327
 QY 143 EEILTEGVAEYHALMDNGHFANDPMRNN---OGYPMNNKPELPLPNCNHAYTGCVATA 198
 Db 328 TOLVREILP-----FNGGRSSIGDMQMDVFIGGNIINDRYIERKNNKSSVNAUL 381
 QY 199 AAOIMKHSMPLOG-----EGSEFDYHAGSLVGNNGTFFGEMYDMINPGNDLON 248
 Db 382 RLPTIRNIS--VOOGASVIDNKNYEGSLKWNNSGILSGSINSEPSFLMG--DNAGKNYQ---436
 QY 249 LQSGVDATVATLMRDVASVMSSEYENG-----SGTYSYVVGALNRNFRY 294
 Db 437 -SISTYDGF-----SLSEYHNDKRVDCGNYNAGSGCEYEST-----473
 QY 295 KRSIQLHVALYTSOEHMDIRGELASGRPVYVYAGNNOISGHAFCVCDGASDGTFF--HFN 352
 Db 474 --SASISITPL-----GWTSTLIGTS---DYT--SSSYKKNHL 504
 QY 353 WGMGVSNCFYK-----LTLSPSLGIGEGIGFTYQET--ITGIEPAKPAEAGTAL 406
 Db 505 SEGYFNQNIYKGRQRMOLSTSTSL---KMMDYNFMPALGIYNSQOLYDKGGYISV 560
 QY 407 PILALKDIAEYKSESGLVNGYSI--YNTGEOSN--LDLGYRLNKADGEVIEVK-----457
 Db 561 TL-----TRASRENSLAGSYNSRKYSISNELFYD--GWTSTNNNDGYEVRREK 612
 QY 458 -----TSSINISWYGEHPSEFSLAPNOLSGINTITLLYRTGTEQEWEPVRAHQ 508
 Db 613 NRHNAGRLSGHIN--NFGDLNCSFSNNKR-----NTNSNNSLIGVYSSPALSD 664
 QY 509 GGY-----VASIVNTTDPNNVYV-----VDNNEKGL---SIVPNS 542
 Db 665 GFYWGSGASGLTKLAGGILIKVKSNDTKKNLVYVTALYGDVSLGSDNAPFVPAALTPAS 724
 QY 543 FVADLNSYEHSTIYV--QFNSDS-----PDEIRTPVAPAL-----STGATADYV 584
 Db 725 LIEDNNNGDKNISVLAPTNMDMFLIPGNVYVEIETKYSVSYIGRGFDKNGTPLSGAHV 784
 QY 585 ISLGVMVAEVPGGSSNYPVWMSKDVLTISEGDTLWYRFSINNORDEMKKIGSVSKPTP 644
 Db 785 INEPVHIDDEDGFS-----F 800
 QY 645 EYTHLFEVGHNOT-----STYLLDMAHNRVLPDTLKNLGLPFNGELVYVFRQTOSS 696
 Db 801 EYV-----GNKTLPLFKGRITTYCQLGKNKV-----HKGIVGVADVICDVNSTSS 846
 RESULT 20
 US-09-769-787-2
 Sequence 2, Application US/09769787
 Publication No. US20030091577A1
 GENERAL INFORMATION:
 APPLICANT: Microbial Technics Limited
 APPLICANT: Gilbert, Christophe FG
 APPLICANT: Hansbro, Philip M
 TITLE OF INVENTION: Proteins
 FILE REFERENCE: PWC/P21128WO
 CURRENT APPLICATION NUMBER: US/09/769,787
 CURRENT FILING DATE: 2001-01-26
 PRIOR APPLICATION NUMBER: GB 9816337.1
 PRIOR FILING DATE: 1998-03-27
 PRIOR APPLICATION NUMBER: US 60/125164
 PRIOR FILING DATE: 1999-03-19

NUMBER OF SEQ ID NOS: 388
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 2
LENGTH: 2233
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-769-787-2

Query Match 2.7%; Score 122; DB 9; Length 2233;
Best Local Similarity 17.0%; Pred. No. 4.4;
Matches 16; Conservative 150; Mismatches 335; Indels 376; Gaps 48;

18 OGHSAVTERALSLRLALROVSLMGOTAV-----SDKSIDVYRQDA 64
3 KGH---WNRRVYSIRKFAVGACSMIGTAVLLGNGINAGESVYVADETLITPAEKPE 59
65 ERGITSQ-----EGSPAYFYVNRGNNEGVALVAADDRIPITLAYSPIG 109
60 EKMIVEEKADKALETKNIVERTEQSEPSSTEALASEKKEDEAVTPKEKY----- 109
110 RFMDMSFNDL--RMLQIYDOEIGLISKAQNLNEILTEGVPAEVHMLMDNGHFAND 167
110 ---SAKPEKAPRISQASNOEPLKEDAKAVTNEV-----NOMIEDR 150
168 PMRNGGYPM---NNKEPLLPNGNHAYTCVATAAQLMRY-----HSMPLQEG 214
151 KYDFNOMWYFKLANSKRAIKPAD-----VSTWKKLDLPYMSIFNDFHESPAONEG 204
215 -----SFD-----YHAGSLVGNMGTGEM-YD-- 236
205 GOLNGGEAMYRTFKLDEKDKKNVRLTFDGYVMDSOVYVNGQLVGHYPNGYNQFSYDIT 264
237 -WNNMGND-----LDNLTQSDVAYATLMDVSASVSMFYENGSTYSYVVGALR 289
265 KYIQKDGRENVIAVHAVNKPSSRWYSGSGLYRDVTLQYTDKVEKNGF--TILPKLE 322
290 NFRYKRSLOLHVRLYTSQEMHDMIR-----GELASG----- 322
323 EQOHGKVERHTVTSKIYNTDOKDEHLVAEYIVERGHAATGLVTRTASRTILKAHSTSLDA 382
323 ---RP-VYVAGNNOISGHAFVC---DGYASD-----GTFHFMGMGVSNGEYKLT 366
383 ILEVERPKLMTVNDPRALYELITRYRDQQLVDAKDLFGYRYHW---TPNEG- 435
367 LLSPTSLGIGGEGITTYOETITGTEPAKTPAAGDALPILAKDIEAEK-----SE 421
436 -----SLNGERIKF-----HGVSLHHDGALGAE-----ENYKAERYRLQOMKE 474
422 SGLNVCYSIYNTGEOSNLDLGYRLNKADGEVIEVKTSSINISWY-----YGE----- 470
475 MGVNSIRITHNPASEQT-LQIAEL-----GLVQEEAFD---TWYGGKRPYDGRFFEKD 526
471 --HPEFSIAPNLOSOGINTIYLLRRTGTGEOWEPVRAHAGGYVNSIKVNTDPNNVYT 528
527 ATPPEA-----RKG-EKV-----SDFDLRTWVER 549
529 VDNNEBKLSIVPNSFYADLNSYHSHTIV-----QPNSDSPDEIRTPVAFAL 575
550 GKNNPALFMSISGNEIGEANGDAHSLATYKRLVKIKDVKTTRYVTMGADKFR---FGN 605
576 STGA---TADVYISLGMVAEYVPGSSNYPVNSKDVLTLSGDDYTLMYRFSINNOKDEM 632
606 GSGGHEKIDELDAVFNFSE-----DNKALRAKHKPLIYGSER----- 646
633 KKGVSVKTPYETHTPLFEVGHQ-----TSTYTLDMANRY 670
647 ---SSATRTGSGSYRPERELKHSNGPERNYEQSDYGNDRVGKATATASWTFDRDNAGY 702
671 LPDFTLKNGLPNGELVYVFRTOSS--SGSLMAADEVYHIKGETFYV-----KP 720
703 AGQFTW--TGTDTIGEPFMHNONQTPVNSYFGIVDTAGIPKHDEYLYOSOMVSVYKKP 760
721 VV-----EGPIPDGSY--RATILHAFVNGOQLYLKGRNRYTVK-I 757

Db 761 MVLHPHMMNENKELASKVADSEKIPVRAVSNASSVEFLNGKS-----LGKTFNKKQT 816
QY 758 VNGTAVEALISSEIRVFPNPARDYVEISAPCIPOETSIILFDLSGRIVKNSLSAGHR 817
Db 817 SDGRTYQEGANANEL-----YLEWKVAQPGTLEAIARDESGREIARDKITTA-GK 866
QY 818 MDYSRLPNGAYILLKVDG 834
Db 867 PAAVRLIKEDHAIADG 883

Search completed: May 29, 2003, 13:28:21
Job time : 35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 13:24:27 ; Search time 26 Seconds

(without alignments)
3116.974 Million cell updates/sec

Title: US-10-030-330-1

Perfect score: 4438
Sequence: 1 MKSFLLAIYMEGIAMQGH.....PNCAYILKVDGYTTKINIVH 843Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Archived: 283224 seqs, 96134422 residues

1 number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR-73.*
2: pirl.*
3: pirl.*
4: pirl.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1196	26.9	886	2 A10890	cysteine proteinase
2	494	11.1	398	2 A37768	streptococcal pyro
3	388	8.7	337	1 BPSOP	streptopain (EC 3.
4	179	4.0	939	2 I41197	ase protein (enter
5	173	3.9	3624	2 AD0835	large repetitive p
6	163.5	3.7	1155	2 H71456	probable pyrolysin
7	157.5	3.5	4936	2 AH2515	hypothetical prote
8	154	3.5	1325	2 A64905	yeak protein - Esc
9	149.5	3.4	1471	2 D72758	hypothetical prote
10	146.5	3.3	1343	2 E90893	hypothetical prote
11	145	3.3	918	2 T02759	hypothetical prote
12	144.5	3.3	1256	1 A43829	muramidase-release
13	144.5	3.3	1343	2 D85724	hypothetical prote
14	143	3.2	3263	2 E82410	hypothetical prote
15	142.5	3.2	940	2 AD1374	internalin protein
16	142.5	3.2	1349	2 A11476	cell surface prote
17	142	3.2	682	2 J00420	beta-1,3-glucanase
18	141.5	3.2	827	2 T20492	hypothetical prote
19	141.5	3.2	1308	2 A90428	serine proteinase,
20	141.5	3.2	1771	2 S76851	hypothetical prote
21	141	3.2	709	2 C64057	IDA-specific metal
22	140	3.2	1348	2 AH1115	cell surface prote
23	140	3.2	1461	2 E90696	hypothetical prote
24	140	3.2	1461	2 A85547	hypothetical prote
25	139.5	3.1	814	1 KXBY	kekin (EC 3.4.21.6
26	139.5	3.1	890	2 A30481	bacteriolysin BCN5 -
27	139	3.1	1417	2 H90670	probable invasins
28	139	3.1	1417	2 D85521	probable adhesin e
29	139	3.1	4199	2 S76412	hypothetical prote

30	138.5	3.1	650	2 S48751	major surface prot
31	138.5	3.1	2902	2 C71953	toxin-like outer m
32	138	3.1	1879	2 S74915	extracellular nucl
33	138	3.1	3283	2 AC1018	large repetitive p
34	137.5	3.1	647	2 A49218	hemagglutinin homo
35	137.5	3.1	649	2 S48752	major surface prot
36	137	3.1	754	2 T14877	hypothetical prote
37	137	3.1	1310	2 AE1749	glycosidase homo
38	137	3.1	1398	2 T28159	pyrolysin (EC 3.4.
39	136.5	3.1	2236	2 AD1018	large repetitive p
40	136.5	3.1	4436	2 E71086	hypothetical prote
41	136	3.1	5188	2 B85547	probable RTX fami
42	136	3.1	5291	2 F90696	hypothetical prote
43	135	3.0	987	2 A64474	hypothetical prote
44	135	3.0	1829	2 AE1664	hypothetical prote
45	135	3.0	1939	2 D97316	probable S-layer p

ALIGNMENTS

RESULT 1

T10890 Cysteine proteinase (EC 3.4.22.-) / hemagglutinin protein - Porphyromonas gingivalis

C:Species: Porphyromonas gingivalis

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999

C:Accession: T10890

R:Madson, T.E.; Clark, V.L.; Kuramitsu, H.K.

J:Infect. Immun. 63:238-247, 1995

A:Title: Revised sequence of the Porphyromonas gingivalis prt cysteine protease/hema

A:Reference number: Z17199; MUID:95105001; PMID:7806362

A:Accession: T10890

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-886 <MAD>

A:Cross-references: EMBL:S75942; NID:9913136; PID:9913137

A:Genetics:

A:Gene: prt

C:Keywords: cysteine proteinase; hemagglutinin; hydrolase

Query Match	Score	DB 2	Length	886
Best Local Similarity	26.9%			
Matches	290	Conservative	155	Mismatches 340; Indels 70; Gaps 23;
QY	1	MKSFLLAIYMEGIAMQGHSAPTKRALSLAR--LALRQVSLRMQOTAVSKISTDIY	58	
DB	19	MKRIFFYLLGLLCLPML-QAGPYTRSKAEQFAKRPRLT-----SSSTASLRMDEV	73	
QY	59	YRQGDARGLTSQEGSPAYFYVANRGNNEGVALVAADDRPTILAYSPIGRFMDMPD	118	
DB	74	YKAAREE-----ALFFVNRKCKDFLLVADDRPEVIGYAFKGFPAARMPD	123	
QY	119	NLRMLOIYDOETGLILSGRAQINELIR--TEGVPAEYHALDNGHFAANDPRMNGY	175	
DB	124	NLRGWLKGYEREMLAVDGKAERPIDIREAKPTDLSSTAPIETGEHMSDPLMDGY	183	
QY	176	PMNKEPRLPENGHAHYCYATAAQMTRHSPVLOGEGSDIYAGSLV---NMSGTFG	232	
DB	184	PFNTLHPLPSGQOAYYCAVATAMGQIMRHKKPEKASGEVDY--DMTGTHHYSGTFG	242	
QY	233	EMVDIMPGNPDLNLTQSOVDAYATLMDVSAVSMSFENGSGTYSYVVGALNNF	292	
DB	243	ETYMVMKPGNIVS-GISPEVKALSTFMRVSVSNMOPADFSGTSTVEPALKEET	301	
QY	293	RYKSLQHLVRLATYSQEMDMIRGELASGPPVYVYAGNOSIGHAFVCDGASDGFHFN	352	
DB	302	HYKSLVYHRSLLPGKEWMDMKELAKNRPVYVYAADSMGHAFCDCGEPGTHFN	361	
QY	353	WGMGVSNGFYKLLTSLPSLGTGEGIGITTYOETITTEPAPKTPAEAGTALPILAK	412	
DB	362	WGMGVSNGFYKLLTSLPSLGTGEGIGITTYOETITTEPAPKTPAEAGTALPILAK	420	
QY	413	DIEAEVSESGLANVGYSTYVMGEGQSNLDGRLAKNDGVEIEKTSINISWGC-YGEH	471	

DB 421 GLO-HNMSDEALDSVYKIKNYSTYAGDVKLAYRLTLPLNGTETNPATVATVIMEDIIGES 479
QY 472 PESFSLAPNOLSGITITITLLYRTGTEOMEPEYRHAOGGVNSIKNTDPNNVYTVYN 531
DB 480 TGNITTPCSOFAGKNTITSLYRTDMDADKELKILMLGVNKEV-TMPAGVAAVY- 536
QY 532 NEKLSIVPNSFVADINSYESTITVQFNSDSDERTPAFALSGATADVISLGVWV 591
DB 537 ADARIYLVKOSLSHNLKAYSDCLSATVYVNGTEERSRTFTALRN-TGRLYFLGRHL 594
QY 592 AEY-PEGSSNYPYVMSKDVLTLESGDYTLMYR-FSINOKDEMKKIGSVYKTPREYTH 648
DB 595 VELHPDEDEGEKSLTITGLKARAGYMLVCTGDMELMEDASMIELASIEVAEHTSTHS 654
QY 649 PLFEVGHNOTSTYTLMAHNRVLDPFTLKL-GLPENGELVYVFRQTOSSSSLMAOET 707
DB 655 SLIVASNPQDLTLVIRANPETLPYTSINEGATSGKEIYVAFKAFSET-FEQAKEEH 713
QY 708 VHIKGETFYVYKFEVGEPP-----IPDGYRATLHAFVNGOOLY-----LGRKNRYT 754
DB 714 MSLAOGETKVLSPELTANSSLYTNAELFPGDIYIYIR-----EGCFMIDIDLFQDYIYR 768
QY 755 VKIVNTAVYALISSE-ETRVPPNARDYVEISAPCIPQESITILFDSLGVKVMKNSL- 811
DB 769 IRLITLSSSDIAGKDVSTIVLYPNADHVHVAIPPTVAGSYLRLFDIOGRMQLSTKIR 828
QY 812 -----SAGH 815
DB 829 ICRVYASRRRTSSSEGH 843

RESULT 2

A37768
streptococcal pyrogenic exotoxin type B precursor - Streptococcus pyogenes
C:Species: Streptococcus pyogenes
C>Date: 31-May-1991 #sequence_revision 31-May-1991 #text_change 22-Jun-1999
C:Accession: A37768
R:Hauser, A.R.; Schlievert, P.M.
J. Bacteriol. 172, 4536-4542, 1990
A:Title: Nucleotide sequence of the streptococcal pyrogenic exotoxin type B gene and related reference number: A37768; MID:90330563; PMID:2198264
A:Accession: A37768
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-398 <HAU>
A:Cross-references: GB:M86905; GB:M35110; NID:9153818; PIDN:AAA26978.1; PID:9153819
superfamily: streptococcal cysteine proteinase
words: exotoxin

Query Match 11.1%; Score 494; DB 2; Length 398;
Best Local Similarity 30.6%; Pred. No. 5.4e-23;
Matches 129; Conservative 73; Mismatches 155; Indels 64; Gaps 13;

QY 6 LLAIYVLEFIAMOG-----HSAPYTKERALSIALRLAROVSLRMGOTAVSDKISIDYY 59
DB 10 LLSLALGFVLANPVADONFARNKEKDSATPFIQSAIKAGARSAD-IKIDKYN 68
QY 60 ROGDAERGITSOEGSPAYFYVANRGNNGYALVADDRIPITLAYSPIGRPMDMPN 119
DB 69 LGCEL-----SGSNMIVYINISG--GFVYVSDKRSPELIGYSTGSPANG-KEN 116
QY 120 LRMWLOIY-----DOEIGLILSGAQLMEELIRTEGVAEYVHALDNGHFANDPRMNO 173
DB 117 IASPMSEYVEQIKENKKLDITYAGTAELINQPV-----VKSLLD-----SKGIHYNO 162
QY 174 GYPMNKEPLRN-----GNHAYTCVATAAQMIRYHSWPIQSGSEFDY----- 218
DB 163 GNPYNLTFTVIEKVGESQFVGOHAATCATATQIMKYHNPYKGLKDYTYTLSSNN 222
QY 219 ----HAGSLVGNMNGTFEGEMYDMIN-MPGNPDLNDLTQSOVDAYATLMDVASVSMSTY 273
DB 223 PYFNHPEKNL--FAAISTROYMNNILPYSGRESNVQKM-AISELMADVGLISYDMXYG 277

QY 274 ENGSGTYVYVGCALNNFRYKRSLOLHVATYTSOEMHDMINGELASGRPVYAGNOS 333
DB 278 PSSGSGASSRYVALKENGTNOSVQINSDSKODMEADIKELSQNPQYGVGV 337
QY 334 IGHAFCVDSYASDGFTHFMWGVNSNGFYKLLSPSTSLGIGEGEIGFTIYOEITGIE 393
DB 338 GGHAFYIDGADGNFHVWMMGVGSDGFFRLDALNPSALGTGAGGAFNGVSAVYIK 397
QY 394 P 394
DB 398 P 398

RESULT 3

streptolysin (EC 3.4.22.10) precursor - Streptococcus pyogenes
N:Alternate names: streptococcal cysteine proteinase; streptococcal peptidase A
C:Species: Streptococcus pyogenes
C>Date: 24-Apr-1984 #sequence_revision 07-Oct-1994 #text_change 07-Oct-1994
C:Accession: S07668; A00978
R:Yonaha, K.; Elliott, S.D.; Liu, T.Y.
J. Protein Chem. 1, 317-334, 1982
A:Title: Primary structure of zymogen of streptococcal proteinase.
A:Reference number: S07668
A:Accession: S07668
A:Molecule type: protein
A:Residues: 1-337 <YON>
R:Yal, J.Y.; Kortt, A.A.; Liu, T.Y.; Elliott, S.D.
J. Biol. Chem. 251, 1955-1959, 1976
A:Title: Primary structure of streptococcal proteinase. III. Isolation of cyanogen br
A:Reference number: A00978; MID:76190087; PMID:1270417
A:Accession: A00978
A:Molecule type: protein
A:Residues: 85-107, 'L', 109-244, 'N', 246-337 <YAL>
A:Note: Trp-298 is at the binding site of the enzyme
C:Comment: This enzyme can remove the activation peptide from the proenzyme.
C:Superfamily: streptococcal cysteine proteinase
C:Keywords: cysteine proteinase; hydrolase; zymogen
F:1-84/Domain: activation peptide #status experimental <PRO>
F:85-337/Product: streptolysin #status experimental <MAT>
F:131,279/Active site: Cys, His #status experimental

Query Match 8.7%; Score 388; DB 1; Length 337;
Best Local Similarity 27.8%; Pred. No. 1.4e-16;
Matches 108; Conservative 63; Mismatches 138; Indels 80; Gaps 12;

QY 26 KERALSIALRLAROVSLRMGOTAVSDKISIDYYROGDAERGITSOEGSPAYFYANG 85
DB 9 KEAKOSAITPFIQSAIKAGARSAD-IKIDRYNLEGEL-----SGSNMIVYINISG 59
QY 86 NNEGVALVADDRIPITLAYSPIGRPMDMPDNLMWLOIYDOEIGLILSGAQLNEBI 145
DB 60 -----MESYVEQIR-----ENKILDTTYAGTAETIKQPV 87
QY 146 LRTGVPVAVHALMDNGHPANDPMRWNOGYPNNNKEPLPN--GNHAYTG-----CYA 196
DB 88 -----VKSLLD-----SKGIHYNGNPYNLITPEYKVPGRGQFVGAATGCHVA 133
QY 197 TAAQIMRHSNPLQEGSEFDY-----HAGSLVGNMNGTFEGEMYDMINMGONPDL 246
DB 134 TATAQIMKHYNPPDKLKNYTYTLSSNPDPFHPKRL--FAAISTROYDMNMI--LPAY 188
QY 247 DNLTOQVD-AVATTLRDVASVSMGTFENGSGTYSVYVVALGRNNFRKRSLOLHVRL 305
DB 189 SGRQSONYMAAISLMADYGISVDMYDYGSSGASRRYORLKENFGTINOSVHOIDKCD 248
QY 306 YTSOEMHDMIRGELASGRPVYAGNOSIGHAFCVDSYASDGFTHFMWGVNSNGFYKLL 365
DB 249 FSKQDMEADIKELSQNPQYVEYEGVKGAGHAFVIDDAGRNFRHYDMGNGVSDGFFRL 308
QY 366 TLSPSTSLGIGEGEIGFTIYOEITGIEP 394

Db 309 DALNPSALGTGAGGAGGNGEYSAVVGIKP 337

RESULT 4

I41197 eae protease (enteropathogenic) - Escherichia coli

C:Species: Escherichia coli

C>Date: 31-May-1996 #sequence_revision 31-May-1996 #text_change 26-Aug-1999

C/Accession: I41197

R:Jersey, A.E.; Yu, J.; Tall, B.D.; Kaper, J.B.

Proc. Natl. Acad. Sci. U.S.A. 87, 7839-7843, 1990

A:Title: A genetic locus of enteropathogenic Escherichia coli necessary for the production of enterotoxin

A:Reference number: I41196; MIMD:91045893; PMID:2172966

A/Accession: I41197

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-939 <RES>

A:Cross-references: GB:M58154; MID:g145852; PIDN:AAA62775.1; PID:g145854

Superfamily: Invasin

Query Match 4.03; Score 179; DB 2; Length 939; Best Local Similarity 20.38; Pred. No. 0.0057; Matches 179; Conservative 119; Mismatches 330; Indels 254; Gaps 48;

Db 9 IYALFGLMGHSAAPYTK-----ERALSRLRLRLROVSLRMGQTAVSDKISID 56

Db 132 LVAAGVA--GHTNKLTKMSPDYTKSMTDKALNTA---AQAASLSGQLSRLNGD 165

Db 57 YVYRQDAERGITSQEGSPAYFYVANRG-----NNEGATVAADDRIP---T 101

Db 186 YA---KDTALGAGNQSQAQMLQHYGTAEVNLSGNFPGSSSL---DFLLPFYDSEK 239

Db 102 ILAASPIGRDMDS-----MPDNLRMLQIYQDEIG---LILSG----- 137

Db 240 MLAFGGVARYIDSRFTANLACQREFLPENMLGYNFIDQDSGDNTRLGIGGEYRDY 299

Db 138 -KAQNEELIRTEG-----VPAEVHALMDNGHNPDP---RRMNGYPMN- 178

Db 300 FRSSVN-GYFRMGSMHSTYKKYDERPANGPDIFRFGYLPSPALGAKLMEDYIDNV 358

Db 179 ---NKEPLRNGHAYTGVAATAQIMRYHSMPLQEGSPDYHAGSLVN---WSGT 230

Db 359 ALFNSKRLQSNPCAAATVG-----VNTPIPLVLMG-IDYRHGT--GNENDLTVSMQ 406

Db 231 FGEWYD--WINMGNPDLDLQSOVDAYATLMADVASVMSFEYENG-----GTY 280

Db 407 FRYQDFPW-----SQIEFOYVNELTLSGSRYDLVQRNNIILEYKODIL 454

Db 281 SYVVAGALRNFRYKSLQHLVRLALYSQ--WHD-MIRGELASGRVYAGNOSIGHA 337

Db 455 SLNIPHDINTEKSTOKIOLIVSKYGLDRIYMDLSLRQ---GGQIOHSGSQAQDYQ 511

Db 338 FVCDGASDGTFFHNMCGVSNNGFYKLLTLLSPTSLIGEGE-----GFTIYOEI 388

Db 512 AILPAYVOG-----SNVYKVTARAYDRNGSNNNVLLITVLSNCGVDQVG 559

Db 389 ITGIEPAKTPAEA-GTDALPILALKDIEAYKSSGSLNGVSYIYNGEESNLDIGRLN 447

Db 560 VTDPYADKTSKADGDEAITTYTA-----TYKKNGVAQANPVSP--N 599

Db 448 KADGEVLEVTSINISWYG-----YGEHPES--FSLAPNQLSQGINTTLLY----- 493

Db 600 IVSGTAV-LTANANNTGSGKATVTLKSDKPGQVYVSAKTAEMTALNANAIVFVDTKA 658

Db 494 --RRTGDEQEPVRAHAGGYVNSIKVNTDP--NNVVYVDNNEKSLIYVNSFVADLNS 549

Db 659 SITEIKADKTTAAVANGODAITTYVVKWKGDKPVSNQVETFTTLGKLS--NSTEKDTNG 716

Db 550 YEHSTIYQV-----SDSPDEIRTP-VAFALSTGATADVDVLSLGVMAEVPBGSS 599

Db 717 YANVTLLSTPGKSLVSARVSDVAVDKAEVEE--FTTLTID--GMEIYVGTGKVG 770

Db 600 NYEVVMSKDVLTLSGQDYTLMYRFSINQNDKWKI-----GSVSVKTPREYTH 648

Db 771 KLPTW-----LQYGVNL--KASGNGKRYTRSANPALASVDASSGQVTLKEKGTITI 822

Db 649 PLFVGHNTSTYTLDMANRVLPDF-----TLKNLG--LPFN-GEIVVYFR-- 652

Db 823 SVIS-SPNDATATYTIATPNSLIYVNSKRYTYNDVAVTCTNFGKLPSSONLENNFKAW 881

Db 693 -----OTSSGSL--WMAOEVVHIKQGEFFYKPPVEGPI 726

Db 882 GAANKYKYSQSTIISWQQTADAKSGVASTYDLVKONPL 923

RESULT 5

AD0835 large repetitive protein [imported] - Salmonella enterica subsp. enterica serovar Typh

C:Species: Salmonella enterica subsp. enterica serovar Typh

A>Note: this species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001

C/Accession: AD0835

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Church

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr

, S.; Moute, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se

A:Reference number: AB0502; PMID:11677608

A/Accession: AD0835

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-3624 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD05867.1; PID:g16503842; GSPDB:GN00176

C:Genetics:

A:Gene: STY2875

Query Match 3.98; Score 173; DB 2; Length 3624; Best Local Similarity 20.38; Pred. No. 0.11; Matches 223; Conservative 102; Mismatches 376; Indels 400; Gaps 54;

Db 14 GIMOGSAPVTERALSLARLRLROVSLRMGQTAVSDKISIDYVYQSGAENGIT----- 69

Db 2545 GIA-AGFTADTPPEAPITTNVVDVGIYTGALANGOVINDAOPTLNGTGAQATVSIY 2603

Db 70 -----SQEGSPAYFYVANRGNGEGATVAADDRIPITLAYS 106

Db 2604 NNGALGTTANNSGNNSFTPTGNTLTCGSHAFATATANNAGTSVSAATVYDTLAPGT 2663

Db 107 PIGFDMDSMPDLRMWLQIYDEIGLISGKQALNEELRT--EGVPAEVHALMDNGHF 164

Db 2664 PSLGLSAD-----GSLSGOAEANSTVTVTLAGVTLTTTA--GSNG-- 2703

Db 165 ANDPMRNGOYPPRANK-----PLP--NGHNAITGCATA 198

Db 2704 ---AWSLTLPKQLEGOLINVTADAGNAGSGLGITAPILPLAARDNITSIDLITSTA 2758

Db 199 AAOIMRYHSMPLQEGSPDYHAGSLVN-----WSGTGEM----- 234

Db 2759 VITQNTNSDGLLVLGLG--NVASVGLNDIAQVETITAEGBGTGVITIDAATGIVLSLS 2817

Db 235 -----YD-----WINMGNPDLDLQSOVDAYATLMADVASVMSFEYENGSGTYS 281

Db 2818 TOEIVQRYDTSIGTW-----TTIVTAVGDFANLLTLGSGVTLMLNGLAGEQYR 2868

Db 282 VYVVGALRNFRKRSQHLVRLALYS--QEMHDMINGELASGRPVYAA-- 330

Db 2869 V-----LTYNTSL--ATGSYSLVDVQTSAG--TISG--PTISTNVMAADTAPT 2915

Db 331 -----NOSIGHAFVCDGYASDGTFFHNMCGVSNNGFYKLLTLLSPTSLIGIG- 376

Db 2916 GTIVATATNANGSTPAGAGV--DILGQYGTILHN-----QGSYITVTLTKPTA--GYGH 2967

Db 377 GEGIGFTIY-----EITGIEPAKTPAEA-GTDALPILALKDIEAY-----KSES 422

Db 2968 KESFTYITONGVSSAAQQLVINLGPAPVPGSVIATDNNAISLV--DTHVSYVNNGSTOS 3026

QY 423 G----LNVYS-----IYNTGGEQ-----SNLD--GYRLN 447
Db 3027 GATVLSVGLGNVLNANLLDDMTNPIIFNVEBAGATTTMLQGTGVSGLVSTFDLTVYFKN 3086
QY 448 KA--DGEVIEVKTSSINISWYGEHPESFSIAPNLSQGINITLTLVYRTGTGEOMEVPR 505
Db 3087 DAIOQYEQFPRQKGINITLLLAGSQSPLTLPLGEGYLFVINTASGISVLNGYTLAISQD 3146
QY 506 HAQGGVMSIKVNT-----TDPNNVV-----527
Db 3147 HTYA--VDSITANTGNVLFDNDVAPFDALLTEVNSVALAATGTEVNLGSLIIDARGN 3204
QY 528 ---TVDNNEGKLSI--VPNSFVADLNSYESTHITVQFNDSDEIRTPVAFALSTGATADD 583
Db 3205 YTYTLAKNGVADSITPPDSFITLAKPAGDDTDASLNT-----TPRARLDIAINDYSD 3257
QY 584 VISL-----GWVNAEY-----PGSSNYPVWMSK-----DVLTLSDG 616
Db 3258 TLSVATLADTAAIMDDSSVGSASWGLGKSGSGSGTFDVAATGYLKGASLVEDVSTL---- 3313
QY 617 YTLWFRFSINNOKEWK-----KIGSVASVPTPEVTHHFLFVGHNGTSTYTLIDMAHN 668
Db 3314 -----ITLGNLNTSMALQENGTVIRNGTVYANIT-----LG---SATVWNLS-- 3354
QY 669 RVLPDFLTKNLGLPEN--GELYVVEFROTQSSGSLMAAE---TVHIKQGETFYKKPVVE 723
Db 3355 -----GLELDAGTYTLNFTGTNTLAGAATTPRVIGTVYDLDNFTSGHNYL- 3402
QY 724 GPIPDG-----YRAITLHAFV-----GOQQLYLK GK 750
Db 3403 GNIEFGSDAAGAMDQNLVNTVRLTSGYNGSAAITLDAANNTSATIOGHYTLIDINLNGA 3462
QY 751 RNYTVKINGTAVAEIESSEERIVFPNPARYVEIASCPIDEMSIILFDGSKRIVMKS 810
Db 3463 YTYTLN--NGVAMSSITSE---VETTYQLDKRIGHT-----DSATITLIDMAPQIVSTNQ 3511
QY 811 ----LSAGHGRNDVSRPLNGA 827
Db 3512 NDVLIGSAYGDTLLYHLNGA 3532

RESULT 6
H71456
Probable pyrolysin (EC 3.4.-.-) homolog PH0310 [similarity] - *Pyrococcus horikoshii*
C:Species: *Pyrococcus horikoshii*
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
Accession: H71456
Varbayski, Y.; Swada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekitani,
Ofuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: H71456
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1155 <RAN>
A:Cross-references: GB:AF000001; NID:g3236128; PIDN:BAA29383.1; PID:g3256700
A:Experimental source: strain O73
A:Note: this accession replaces an interim accession for a sequence replaced by Genbank
C:Genetic8:
A:Gene: PH0310
C:Keywords: hydrolase

Query Match 3.7% Score 163.5; DB 2; Length 1155;
Best Local Similarity 19.5%; Pred. No. 0.071;
Matches 216; Conservative 129; Mismatches 351; Indels 409; Gaps 60;

QY 15 IAMGSHAPYKKEALSLARLROVSLRMGOT-----AVSDKISIDYVR 60
Db 54 LRLTGHTIYPWDEAF---REALTKYAPYIGRTLPLSAGNSGNOLPBRVETVETLPPVGN 110
QY 61 QGDAERGI-----TSQEE-GSPAFVY-VANRGNNEGVALVAAD 96

Db	111	OGHVSSCGMSSSTYYWMTYMLMMWRNNPHPTSPEDIMNPTFYTLINGADGSGWMMAD	170
QY	97	DRIPILAVSPIGRDMDM-----PDNLR-WMLDIOEIGLISGAQNLIELIRTEG	150
Db	171	NVISTIGAV-PLNAPPLVYLGRRGDPENYAWWPMPLTOMMAAPVNSGYDM-----	220
QY	151	VPAEYHALMDNHFANDMPMR-----NOGYPMNNKEPLLPNGNHAATGCVATAAQIM--	203
Db	221	---YMQWMDPYNYTPEGMMWTLVLDNGTQMNLYKGLLSSGYVQTMVLVPSREFLNR	277
QY	204	-----RHS---WPLQGBSGFDYHAGSLVGKMSGTFGEYMDIMNPGMD	245
Db	278	PEDFLNMNLFYEVKSYSDLYMTSDG---YKY-----W--TVEGLDLMVSSAYNET	324
QY	246	LDNLTQSOVDVATLMRDVSAS-----VSMSEYEN--GSGTY-----	280
Db	325	FER---DPSIYIELMKEVLYNDEGVSLTDTIPVAEKINATFYRFRPNQGTMDAEAFY	380
QY	281	-SVYVYVGLRNNFRYKRSLOLHVRLYTSQEM--HDMIRGE-LASGRPVYAGNNO---	332
Db	381	LSAY---SLKGEMVYVNNAFVDLYAL-ANFEMLMLEYIEGGRVSSGRALAYVFNMYWAM	436
QY	333	SIGHAFVODGYASDGT-----FHRNMG--WGVSGNGFVKULLLSPTSIGIGSEIG	381
Db	437	AGGHAVTITIGDNNVTPTDGGGALLMVAWSGTDWG--DNGEMKSYEART-----	485
QY	382	FTIYO-EIITGIEPAKTPAEGTALPITALKD-----IEAEYK---SESGLVGY-	428
Db	486	YTWYDYPILISVR-----GLIDPLFISYDEAFVVPRAKAKYKPLMTVYGINHBYR	537
QY	429	---SIYNTGEEQSNDLQYRLNKADEVIEVTKSSINISWY-----	466
Db	538	GEVVDGIYNS-----TTYELLASSGILAGVKVNN-TIWFENHYLDFWIDYISDKEN	588
QY	467	-----GYGHEPESFSLA-----	479
Db	589	ALISNLSLLESYLPOAHPPESPMAFDVQSJLDLYVYMEDKNSPTFETVYVQKLLP	648
QY	480	NOLSOGINTITLL-----YRRTGEOMEVYRHAOGYV--SIKVTTD--PNNVVT	528
Db	649	DNFGELNPLNLLISIGGEYKVLGSLBGN-VSIPDGGYVTPLEIPVYSYDVAAPVNVQ	707
QY	529	VDN--NEGKLSIYP-----NSFVADL---NSYERSTTV	557
Db	708	XGNEVWSIYIIPLEDARIYVIGNESYPLTAEGGIYFIANSIAELKLIHAQTYNTVYT	767
QY	558	QFNSD-----SPDE-----IRTPVAFALSTGATADVI	585
Db	768	YPNGKEVALPERTVITASPTVSIISPEETINETVPIPVKAVKDAVEIVANTAKGEVL	827
QY	586	SLGWMVAPEVGSSNYPYVMSKDVLTLSGDTYTLMRISINNQDEMKKISSVSKPTTE	645
Db	828	ELTY-----NESSQY---YTGTLINISDGAFTL---VTVAVDTMNTNTVAKV----	867
QY	646	YTHPLFEVGHQTSYTYLDMAHNRRLPDTLKNLCLPENGELVYVFRTOGSSGSLMAAQ	705
Db	868	--HPIYAKAGAVT--TYIYENA-----IVNCGYGN--YIYVGNKKIVYANVTSE	912
QY	706	ETVHIK-----QGETFYVKPVVEGPIPDGYSRATLHAFFVNGQOOLYIK-----	751
Db	913	GTYEVKAVYVNNVPNSIFVNSAISIES-VATGEANATLVAGMNASYEATTSVGEPEKEDKL	971
QY	752	NYTYKIV-----NGTAVEA-----JESSEIRVFPNP---ADAYE--ISA	787
Db	972	LYPTATLADAVELGNGAVVIALRDVNISKIYLIKNGOKIOLTTNESDPRIAYVYONGIYF	1031
QY	788	PCIPQETSIIILFDSGKIVMKNSLS	812
Db	1032	VVIKQDPRIVAYGYAKPAPRRKYS	1056

Db 705 GHTVFNAGNTYSGKTLVNDGLLTIASHTAGVTGKG--SSEVTIANPG-----TLD 753
Oy 609 VL--TISEGDTLWYRS-----INNOCKDEKKIGSVKPTETETPLFEVGHONSTYT 662
Db 754 ILASTSAGDTYTLTNALKGDLKRVOLSSDDKMGFTGATGTEFA---GVALOKDSTFT 809
Oy 663 LDMAHNRVLPDFTLKNLGLPFNGELVYVFRQTOSSGSLMAAOETVHIKOGETFFV---- 718
Db 810 LERONTAALTLHMLQS-----DSENTTSVKVGEQSLIGGL-----AMNGCIIITDID 856
Oy 719 --KPVYEGPIPDOSTYATLHAFVNGOOLYLKGRNITYKIVNGTVAEISESEETVRP 776
Db 857 PAATLLEGTI-----SVDTLVVAGADYTWKG-RNYQ---VNGTGDVLID-----VP 898
Oy 777 NPARD 781
Db 899 KPMND 903

CT 9
58
Hypothetical protein APE0057 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: D72758
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A:Reference number: A72450, MIMD:99310359; PMID:10382966
A:Accession: D72758
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1471 <KAN>
A:Cross-references: DDBJ:AP000058; NID:95103388; PIDN:BAA78966.1; PID:d1042742; PID:g510
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0057

Query Match 3.4%; Score 149.5; DB 2; Length 1471;
Best Local Similarity 18.3%; Pred. No. 0.76;
Matches 192; Conservative 158; Mismatches 352; Indels 345; Gaps 48;

Oy 13 FGIAMOGHSAPYTKERALSRLARLAKQVSLRMGQTAVSDKISIDYVYRQGDAREGI--T 69
Db 41 YTIAYGGAIVYAVAGNVAVASYSSENSVELVTG---DGLATVYVYAREPRVYVSLXS 95
Oy 70 SOEGSPAYFYUAN-RGNNEGVALVADDRIPITLAYSPIGRFDMDSMDNLRMLQITD 128
Db 96 GGEKVAQASFEIIOULSGESDPRPDIVDDSLAVSTLTSPPSGVEILLAPFIPSSD 155
Oy 129 OEIGLLISGAQNLNEILRTGEVRAVNAIMDNGHAFANDPMWNOGYPMNNKEPILPN 188
Db 156 GAAVYVUTSIVYDRV---SISVDNTKRTREYGVADALLIGASGVREPKADITYQG- 210
Oy 189 HAYTGCVALTA-AAQIKRYTSPLOGSGFDYHAGSLVGMWSTFGEMTDMWMPGNPDID 247
Db 211 -AVLRFKASGDSGFVVKLSFNLITDGLVLRGLGTG-----LELE 251
Oy 248 NLTOGQ-----VDAYATLMRDVSAVSMSFENGSGTYSYVUGALRNPNFRKRS 298
Db 252 YLTIDASGRVKLRLPSTISL-GDALATGNTISLHNVEGFLNVQVNSDPRKI----- 304
Oy 299 QLVHVALYTSQEMHDMIRGELASGR---PVUYAGNNOSIGAFVCDGYASDSTFPHNMG 355
Db 305 -LAIVLELISGRPMRPGI---ELVSKHNHVRLY-----IYEGRASQMEFLIRIGV 348
Oy 356 GGVS---NGFYKLTLSPLSLGIGGEGIGFTTYOETITGIERAKPRAEGDRLPLAK 412
Db 349 GVLALEPISILKLT--PLEVLEGG---SYFYKQRLAVALVYASTNP---TEDRLDLAP 400
Oy 413 DIEAEYKSRSGNLNVY-SIYNTGEOSNLDLGYR---LNKADGEVIEVKTSINISWVG 467

Db 401 SVEVSSKSGSG---GVFAIAGIWLAEADMSIGNESSLVLTALSYAKLDVDFAINSSTII 457
Oy 468 YGEHPS-FSLAPNLSQGINITTL-----LYRTGTQMEPVHQAQ-GYVN----- 513
Db 458 IGRPPEGLPVAPPEADTILNTVDLNVAEAEIARDAIKADIEFVASSGGITMGAVLA 517
Oy 514 -----SIKYN-----TDPNNVY---TVDN-----EGKLSIVNSFVA 545
Db 518 EAONLSKNGDAVSAIVASASSILGDIYLSNNIEAVFSTTILTSPIETRISLGRNTSLT 577
Oy 546 DLNYSHTTIVQ-----FNSDPDIRPVAPALSTGA----- 579
Db 578 TYSLSHSPLIYASGTNGSSTFGASVSTKLGLFNLTPD-----LTIYSSVLELEA 632
Oy 580 -----TADVISLGMVAEPGGSSNPVWMSKDVLTLSGDTYL 619
Db 633 YKTRGSITLMREGKASIELNTPDSQSL---LAIWGS-----ITLEDSSFK- 677
Oy 620 WYFESINQKDE-----WKRI-----GSYS-----VKTPPEYH 648
Db 678 --KFTYKTRDRLTIFMKGVAVYERASFEVSYMSKLRKGYDGLSASLSMTAPEGEI 735
Oy 649 PL-----FEVGNH-----OTSTYTLMAHNRVLPDTLKNL-----GLPFNGE 686
Db 736 TLEDTLISGVNFRVDESEAMPYRFTSNVYEGGSNSAI---TVKSVETIISTYINAT 792
Oy 687 LVVYVFRQTOSSGSLMAAOETVHIKOGETFFVYKPVYEGPIPDOSTYATLHAFVNGO---- 742
Db 793 YTVIVSEIASISGSSIQG-TLNTIQASNLV-----LQDNTLGEVNTILSSQTSV 843
Oy 743 --QOYLKGRNITYKIVNGTVAEAEISESEETVRPAPADVEIASPCIPQETSIILFD 800
Db 844 EVKRLTVECSQITLIDGVG-----YVESASGGSTIALIQFN 882
Oy 801 -----LSGKITYAKNLSLASHGKRDVSRLEPNAYI----- 829
Db 883 GGVTYLTGTV-----EGELEYS-LSNGARLMSMLEVKGGLRIYVGEDYVEGSA 932
Oy 830 -----LKVQDYTKINIV 842
Db 933 SLVEICIASASATSLSEVSGLTADLNIV 959

RESULT 10
E90893
Hypothetical protein Ecs2117 [imported] - Escherichia coli (strain 0157:H7, substrain
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: E90893
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shimagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and g
A:Reference number: A99629; MIMD:21156231; PMID:11258796
A:Accession: E90893
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1343 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA835540.1; PID:g13361583; GSPDB:GN00154
A:Experimental source: strain 0157:H7, substrain RIMD 050952
C:Genetics:
A:Gene: Ecs2117

Query Match 3.3%; Score 146.5; DB 2; Length 1343;
Best Local Similarity 21.7%; Pred. No. 1;
Matches 155; Conservative 84; Mismatches 244; Indels 231; Gaps 40;

Oy 161 NGHFANDPMWNOGYPMNNKEPRL---PNCNHAATGCVATAAADIIMRYHSPLOGEGSFD 217
Db 295 NGYFNGCAVNISNNGLINNKEYSIVGVQDGSBGVNVNTDK----- 334
Oy 218 YHAGSLVGNMS--GTFGE--HTDWIMNPGNFDLNLIOQVDAVATILMRDVSASVSMSTY 273

Db 335 -----GHNNEIGT-GEARFYITIGDAGGELVNSREGKVD-----SGITIGMK 377
Qy 274 ENGSGYV-----YVVALKNNERYKRSLOLVH--RALYTSQEHMDIRGELASGRPVY 327
Db 378 EETGNTLVKDKNSVITNLGTNLGYDGHGEMINISNEGLVYVNSGSSIGYGTGVKGSIT 437
Qy 328 AGNNOSIG-HAFVODGYASDGTFFHNMG-----WGVSNGFKTLTSTST----- 372
Db 438 TGGIWEKNKVVYTTIGVAGVGNLNI SDGKRFVSONITFLDGKASGIGTLNMDTSSFD 497
Qy 373 -----LGIGEGE-----GFTIYOE-----ITGIEPAKTPAEAGTALPILAKIDEAEYS 420
Db 498 VGINVNGSGGIYVNSGATLNSGCTGFIIGNASGKIIVNISTDLNKLSTSTNADL-- 555
Qy 421 ESLGNGYSIYNGEOSNLDLGRKLKADGEVLEKTSI-----NISWYGEHP 472
Db 556 ---LVG--VLGTG-----LNIITGGIYKADTQIALNDKSGKDVYDQNSLL 600
Qy 473 ESFLAPNLSOGINTITLLYRT-----GTEOW-----EPV-----RHAGGYV 512
601 ETRNMYG--TSGITLITLNSGTLNVEGGEVYLGVEPAVGLNLIGAHEAADAAGFI 658
Qy 513 -NSIKV-----NTDPPN--NYYVTVDNNEGKLSIVPNSFVADLNSYEST 554
Db 659 TNAKVEFGSGEGEYVENHTNNSDAGYQVMDLITGDKDKYI-----HDA 704
Qy 555 ITVOFNSDPELRTPAFALSTGA--TADVISLGVMAEV-----PGSSNPVYVMSKD 608
Db 705 GHTVFNAGNTYSGTILVNDLITLASHADGVGMG--SSSEVTLASPG-----TLD 753
Qy 609 VL--TISEGDTLWYRS-----INNOKEMKIGSVKPTETETHLFEVGHNOTSTYT 662
Db 754 ILASTNAGTYTLNLAKGGLMARVQLSSDPMKGFTHATGERA---GAQLKSTFT 809
Qy 663 LD-----MAHNRVLPF--TLKNLGLPENGELVYVROTQSSSGSLMAAETVHIKOG 713
Db 810 LERDNTALTHAMLOSIEMTISVNG-----EOSIGL-----AMNG 847
Qy 714 ETVYV-----KPVESPIPDGSYRATLHAFVNGQOOLYLKGRNTYVKIYNGT 761
Db 848 GTLFTDTPAATLAEGYI-----SVDTLVYGASDYTWKG--RNYQ--VNGT 890

RESULT 11
T02759
hypothetical protein - Acinetobacter sp. (strain ADP1) (fragment)
Species: Acinetobacter sp.
Title: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999
Accession: T02759
R:Segura, A.; Ornstom, N.L.
Submitted to the EMBL data library, June 1997
A:Description: P2R153.
A:Reference number: Z14724
A:Accession: T02759
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-918 <SEC>
A:Cross-references: EMBL:AF011339; NID:g2286203; PIDN:AA027114.1; PID:g2286204
A:Experimental source: strain ADP1

Query Match 3.38; Score 145; DB 2; Length 918;
Best Local Similarity 18.08; Pred. No. 0.68; Mismatches 282; Indels 344; Gaps 45;
Matches 165; Conservative 127; Mismatches 282; Indels 344; Gaps 45;

Qy 94 AADRLPTILAYSPIGRFDMSMDNLRMLQIYDQIGLILSKAOLNEELITRECVPA 153
Db 97 ATDDPTPLVGTGVGAV-----VLSVDGGAAGVAVVDSNG--- 134
Qy 154 EVHALNDNGHFAANDPMKRNOCITPMNKEPILPNGNHAATGCVAATAAQIMRYHSWPLQGE 213
Db 135 -----NMSYTLPSQSEGTSHYSTGTASNAAG-----TQGT 163

Qy 214 GSPDYHAGSLVGNMSTGTEGEMTWIMPNPNDLNLQ--SQVDAATLMDVSASVMS 271
Db 164 AFTLLTIDTAPD--VPISIGOVNDVGLIOGP---LTQASTDPTPLT----- 207
Qy 272 FYENGSGYSVYVVALKNNERYKRSLOLVH--RALYTSQEHMDIRGELASGRPVYAGNN 331
Db 208 ---TGT-----GATPG-----DVAK-----VYQND 225
Qy 332 QSIHAFVCDGYASDGTFFHNMGCVSNGEYKLTLLSPYSLG-IGEGIGFTIYOEIT 390
Db 226 -LVGSYVY---GADGMNNTYIPAPGLESGHDSVATDPVGNESGSDPFLTVDLTA 280
Qy 391 GIEPAK-----PAEGTALPILAKD-----IEEYSESG-----LNV 426
Db 281 PPAEGTENGSGSEITGSAEAGS---VVKVKGANGVLSATADSSGKTYTTLADPLND 336
Qy 427 GYSIYNT---GEE-----OSNLDLGRKNADEVEYKTSI 461
Db 337 AETVKITATKAGNESQATDLTAPNVIIDADNIVEAKVDPTTYVTPVSETIINETSIL 396
Qy 462 NISWYGEHPESFLAPNLSOGINTITL-----LYRRTGEQWEPVRAHQ 508
Db 397 NI---GSTYPGYPTVQODIADAVISVSTGSLINLFDAEMKLYKODDGSWELIADNQ 453
Qy 509 G-----GYVNSIKVN----- 518
Db 454 SPGLLDLILGFGQTTKYTAGELPGRNTRFDTGSLIGLSTIKADLQTLQNTAANPV 513
Qy 519 ---TTDPNNVYVTVDNNEGKLSIVPNSFVADLNSY-----EHSITVOF----- 559
Db 514 GDITSKGNVITTDADATNGOQVTSQKRYTAVNOGTVAADGTTITVGEHGLITKADGSY 573
Qy 560 ---NSD-----SPDEIRTPYAFALSTGA--TADVISLGVMAEVPGSSNPVYVMSKD 608
Db 574 KYTPNSDVTYIGKIDFTFNITTTDA-STGKSPTAKILIOI-----GTNSDLDLTWPN 624
Qy 609 VLITSEDDTYLWTFNFSINNOKEMKIGSVKPTETETHP-----LFEVGHNOTSTY 661
Db 625 ---NPEADAT---SVAINDE-DVIGIGATNTEETTAGSINESWILGIGSQTYVSO 675
Qy 662 TLDMANHRV-----LPDTLKNLGLPENGELVYVROTQSSSGSLMAAETVHIKOG 715
Db 676 TITIAQNLGAVELGISTSLILG---GAASVQFQKLVNNT---WTVDTISASSLAD 728
Qy 716 FVYKPVYEGPIPDGSYRA--TLHAFVNGQOOLYLKGRNTYVKIYNGTAVPAISSEIR 773
Db 729 LI-----GLFPNGTGVYDLEA---QYRYIILNNG--LGVAAGVSVSSEVTSDDL 777
Qy 774 VEPAPADYVEISAPCIQERSITLFDLSGKIVK-----NSLSAGGRMDVSR----- 823
Db 778 SYTTSREYVQ-----GNVLTDTGTAGVCRVASHYTVDAISHDGLTYTTVTGTGI 827
Qy 824 ---PNCAYILKVDG-YT 836
Db 828 SILTGHITVIRKSDGTYT 845

RESULT 12
A43829
muramidase-released protein precursor - Streptococcus suis (type 2, strain D282)
C:Species: Streptococcus suis
C:Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 16-Jul-1999
C:Accession: A43829; S21549
R:Smith, H.E.; Vecht, U.; Gielkens, A.L.; Smits, M.A.
Infect. Immun. 60, 2361-2367, 1992
A:Title: Cloning and nucleotide sequence of the gene encoding the 136-kDa dalton surf
A:Reference number: A43829; MUID:92267650; PMID:1587602
A:Accession: A43829
A:Molecule type: DNA
A:Residues: 1-1256 <SMI>
A:Cross-references: EMBL:X64450; NID:g47550; PIDN:CAA54781.1; PID:g47551
C:Superfamily: muramidase-released protein

C.Keywords: tandem repeat; transmembrane protein
 F:1-47/Domain: signal sequence #status predicted <SIG>
 F:48-1256/Product: mutamidase released protein #status predicted <MAT>
 F:663-681/Domain: small repeat <RP1>
 F:839-861/Domain: small repeat <RP2>
 F:953-1006/Domain: large repeat <RP1>
 F:1064-1084/Domain: small repeat <RP3>
 F:1089-1142/Domain: large repeat <RP3>
 F:1143-1195/Domain: large repeat <RP3>
 F:1232-1248/Domain: transmembrane #status predicted <TM1>

Query Match 3.3%; Score 144.5; DB 1; Length 1356;
 Best Local Similarity 18.7%; Pred. No. 1.2;
 Matches 214; Conservative 121; Mismatches 374; Indels 433; Gaps 60;

21 SAPVTERALSLARALROVSLR-----MGOTAVSOKISIDYVROGDAREGRTS----- 70
 83 SAVATTSEVEKAAVLEQVTSSEPLAGLQKELATEDATLAKAIEDAQTKLAARAI 142
 71 -----OEGSPAYFYVNR--GNN-EGYA---LVAADRIPTILAVSPIGRFMD 114
 143 LADEATVEQVAVKVAANALGNELQKVTVDGLTALADTV----- 187
 115 SMPNLMMLOIYQDEGLISGAQNEELLREGEVPAEYVHL-MDNGHPANDPMRWQ 173
 188 -APDTTASTLVGDGE-GTLIDSTTTATPSMAEPNGAIAIAPHLRTODGAKATSEPMW-- 243
 174 GYPNNKEPLLPGNNAH---YTCGVATAAQAQIMRHSMPLOGESFPDHYAGSLVGNMST 230
 244 -YFESYDLXSYNNMASTYK-----AEVDATIRSLDNDSTTAVLAELVSRITGD 236
 231 FGEYVDMINMP-----NDLNLQSOVD-AVATLMRVASVSKSFYENGSGYVSY 283
 297 VLEKY--TIPGSESVTFSHPTKYVANNSTITVYDLSANPGALKEFSAN-DOVYSFI 353
 284 VVGLAR-NMRY-----KRSLOL-----HVRALYTSQEW---HMINGELA 320
 354 IYVAYQINTRYTESGVATATGLQTIGQVTPSSVR-VFQYIDVATTTAAVGGYP 412
 321 SGRRPVYAG-----NNOSIGHAFVCDGASDGFHFMNGMGV--SNGF 362
 413 KG--TVYLAGVQKDTQVQKVIKEIVENDQ--VLKFYIDLPYTKGEVDMGCTITGE 466
 363 YKLTLSPTSLGSGISGIFTIYQELIIG-----IEPAKT----- 397
 467 IELLTSPITTKVG-----TIYDYNISKITAPFTIDPKVMYKESONEOGSKYV 520
 398 -----PAAGT----- 403
 521 IAOMSGDETTKGIYKIYATQVWTKLGTNEMGWEDYDDOAGIKFNKGFMPAGVNT 580
 404 --DALPILAK-----DIAEYKSEGLNAGVSIYNTGEE---OSNLDLGR 445
 581 LKNATATATVETTYIKESSKYGDVLEVEYDIDKQIVNSVYDTPKSALETENIDVDR 640
 446 --LNKADGEV-----IEVKTSSINISM----- 465
 641 PASLVAADTVFYFKKVSADSACTTGTVAGTTVAKYVTEKAGSVANVNDINGAVIKAP 700
 466 -----YGT-----GEHP-----ES 474
 701 VSEDKAPKPGVNDTDLDQKLASITFEKREYKLVAPGDPVGVKGNLIEVGNNITAG 760
 475 FSLANQLSOGIN-TITLLYRTGTBQWEPVRAHAGGVNISKV-----TTDP 522
 761 IDPTTGKLEAGVNAKEVTVYRAVTSYVYVNYKDTGINKPDETVDSPADVATYTDK 820
 533 N--NVVATVDNN-----EGKLSI-----VPSFVADLNSYHS 553
 821 KPNELITDGSRYVLPKTDGEENGKVIETGITTYTYOYKAMTIP--IPNVETPRP 878
 554 TITVOPNSDPELTPVAFALSTGATADVYISLQWAAVEVGGSSNPVYWSKDVLTLS 613

879 KVPPEPDPPEPDEPDE-----TTPGTNGEVPNIPY-----VPG-----YTPVDPKNDTPRK 925
 614 EGDYTLMYRPSJNNOKDEMKKISV-----SVKPTREYTHLFEVGNQSTVYLDMA 666
 926 PID-----PNDPGK--GYVPTPEPNPVDPIPIPV-PVKVYVNNH-----VDEE 966
 667 HNRVLP--DPTLKNLGLP---ENGELVYVFPROTQSSGSLMAQETVHIKGE----- 714
 967 GNDIAPQEEGTPKNSIPGEYEFKTV-----TDEGNTHTHYKKPVEKNGTVVNVYT 1021
 715 ---TFYKRPVVEGPI-PDGS-YRATLHAFVNGQQLXGKNRYTKYNGTAVALEISS 769
 1022 EDDTVAKEPYTDPTPTSEGTPIYDT---DNKPKITFKGE-EYELVRVDGTENGKVEG 1076
 770 EEI-----RVFNPARDYV-----EISAPCIQPE-----TSILFDLSGKIYMKNSLSA 813
 1077 ETVVTVYVRKVEVTPAKKVVYTNHVEDGNPNVAAQDEBTKPNKSIPIGEFTGKVTDEDCNT 1136
 814 GH 815
 1137 TH 1138

RESULT 13

D85724
 hypothetical protein ydek [imported] - Escherichia coli (strain 0157:H7, substrain ED
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 R:Accession: D85724
 R:Name, N.T.: Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Llam, A.; Dimalanta, E.; Potamousis, K.; Apoda
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli 0157:H7.
 A:Reference number: AB5480; MUID:21074933; PMID:11206551
 A:Accession: D85724
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1343 <STO>
 A:Cross-references: GB:AE005174; NID:g1251519; PIDN:AA656256.1; GSPDB:GN00145; OMGP:
 A:Experimental source: strain 0157:H7, substrain EDL933
 A:Genetics:
 A:Gene: ydek

Query Match 3.3%; Score 144.5; DB.2; Length 1343;
 Best Local Similarity 21.7%; Pred. No. 1.3;
 Matches 155; Conservative 83; Mismatches 245; Indels 231; Gaps 40;

161 NGHNPDMRNNGIYNNKEPLL---PNGNHAYTCVATAAQAQIMRHSMPLOGESFD 217
 295 NGYFNGAVNLSNNGLINNKEYSLVGVDGSHGVNVYTDK----- 334
 218 YHAGSLVGNWS--GTFGE--MYDMINMGNDLNLQSOVDATATLMRDVSAVSMSFY 273
 335 -----GHMNFLET-GEAFRIYIIGAGXGELVNSREGKVD-----SGITTGMR 377
 274 ENGSGTYSV---YVGLARNNFRYKRSLOLHV--RALYTSQEWHDIMRGLASGRPVY 327
 378 ETGNGNLTKVKQNSVITMLGTNLGVDGHEMNINNEGLVYNSNGSSSLGGETGKASTY 437
 328 AGNNSIG-HAFVCDGASDGFHFMNG-----WGVNSGFTKLLSPTS--- 372
 438 TGGIWEVKNKYTTIGVAGVGNLNLISDGKRVSONITFLGPKASGIGITLMLDATSSFD 497
 373 ---LGIGEGEIT---GFTIYQE---ITIGIEPAKTPAEGCTDAILLAKIDEAYKS 420
 498 VGIVNGNFGSIVAVNSGAILNLSGYFTIGNMSGKQIIVINSDSLMNLKTSSTNAOL- 555
 421 ESSLNAGSIYNTGEOSNLDLGLYRLNKADGEVLEVTSSTI-----NISWYGEHP 472
 556 ---LVQV--VLGTGE-----LNTTGGIYKADQDIALNDSKSGVRRDQGNLSL 600
 473 ESFLANQLSOGINTITLLYRRT---GTEQW---EPV-----RHAOGGYV 512

Db 601 ETEFMVVG--TSGTGLTLTNSGTLNVEGGEVYLGVEPAVGLNLGAGXAADAGFI 658
QY 513 -NSIKV-----NTDPN---NVVTVDNNEKRLSTVPSFADLNSYHST 554
Db 659 TNAIKVEEGSEGEVFNHTNNSDAGYQVMDLITGDKDKVI-----HDA 704
QY 555 ITVOFNSDSPDEIRTPAFALSTGA--TADVISLGVMAAEV---PGSSNRYVWMSKD 608
Db 705 GHVFNAGNTYSKRLVNDLTLTASHTADGVTGMG--SSEVTLASG-----TLD 753
QY 609 VL--TLSEGDYTLMYRES---INNOKDEKKIGSVKPTPEYTHPLFEVGHNOTSTYT 662
Db 754 ILASTNSAGDYTLTNALMGDLKMRVOLSSDKMGFFHATGEPA---GVAQLKDSFTL 809
QY 663 LD-----MAHNFVLPDF---TLKNGLPENGELVYVFRQTQSSGSLMAAQEVHLKOG 713
Db 810 LEEDNTAALTHAMQSIENITTSYVNG-----EOSIGL-----AMNG 847
QY 714 ETEFVY-----KPEVEGPIDPGSYRATLHAFVNGQOOLYLKGRNRYVKIYNGT 761
Db 848 GTLIFPDIDIPATLABEYI-----SVDTLVVGSADYTWKG--RNYQ---VNGT 890

RESULT 14

E82410
hypothetical protein VCA0849 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: E82410
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301
A:Accession: E82410
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3263 <HEI>
A:Cross-references: GB:AE004412; GB:AE003853; NID:99658269; PIDN:AAF96747.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA0849
A:Map position: 2

Query Match

3.28; Score 143; DB 2; Length 3263;

Best Local Similarity 20.08; Pred. No. 6.6;

Matches 170; Conservative 110; Mismatches 288; Indels 284; Gaps 45;

Db 15 IAMQGHSAPTYKERALSLARLROVSLRMGTAV-----SDKISIDVYR 60
QY 334 VALNDGSAFVWDE---PILSLNENDLPAGSDADPLTVSGQDTOLGSDQVA--SYQID 388
QY 61 QGDAAE--RGITSGEE---GSPAYF---YVANGNNEGATL----- 92
Db 389 PSTANPIAGLTSGDGAATILGEPLIDGNRYQAATAGRDIFQLTLNADSGVQFVLTLD 448
QY 93 -VAADRIPTITLAYSPIGRDMD---SMPNLKRMTOIQOEGLILSKAQLNEELRT 148
Db 449 HAAGSDALTTISL---PIVAIDYDNDSSAPGNLN--IEIODKPIITIGAEOLTVAEQTLDT 503
QY 149 EGVPAEVAHLMNDGHFAN-----DPMR-----WNGGYPMNNNEPILPNCHNA 190
Db 504 GSIGGCA-SLVADONFTTGGSDGVSYSLDLSLDVAGITSGVAATVLSGVANNGNYT 562
QY 191 YTCGVATAAQAQIMKRYHSMPLQEGSPDYHAGSLVGNMSGTGEMYDWINNPGNDP----- 245
Db 563 YT---ATAG-----GEPEV---TILLNDGSYR-----FTLQSLDHALNS 597
QY 246 ---LDNLTQSDVATATLMDVASASMSFEYENGSGTYSYVVGALNNRERYKSLQHLV 302
Db 598 DELLVNT-----VVAIDFDGDTASITLPTVTKDKDPY-----FTNYSVLNVEH 641

QY 303 RALYTSQEHMDMIGEL--ASGRPYVYAGNNO-----SIGHAFVCDGYA 344
Db 642 NDL---FQSDVYIKPELITASGQFELVQGSDBVASFTLDSSVNPVQGLTSGNAVTLSPV 698
QY 345 SDGTFHNMGM---GGV-----SNGEYKLLSPSLGIGEGIGETTYQELITGI 392
Db 699 DDG--HGNLTATAMAGAVFTLTLNTDGYFSFTLAPVHALNSNDLTIN--FCVIATDF 755
QY 393 EPATRPAGADLAPILALNDIEAEYKSEGLNAGYIYNTGEQSNLDGLRLKRDGE 452
Db 756 D---GDSISIVPV--KINDDKRYFTNVOGLYV-----HENDLPQ---SDYDKE 797
QY 453 VIEVKTSSINISWYGEHEPESFELAPNOLSGINTITLLRYRTGTQWEPVRAQGGYV 512
Db 798 PVIYN-----GQF-----QVQAGDYV-----ASRALDSSV 823
QY 513 NSIKVNTDPNNVYT---VDNNEKRLSTVPSFADLNSYHSTITVOFNSDSPDEIR 568
Db 824 NPVQGLTS--NGAVVTLAPVDDGNGNL-----TYTAMAGSVYFTLTLNSDGYSPFLA 876
QY 569 TPAVAFALSTGATADVISLGM--VMAEVPGGSS-----NYPVMSKDVLTLEGD 616
Db 877 APVEHALNS---DSLTLNFKYATDFDGFASIVLPVTVLDQPSVISNAQALSVNEDD 931
QY 617 YTLMYRESINNOKDEMKIGSVSKPTPEYTHPLFEVGHNOTSTYTLN-MAHNFVLPDF 675
Db 932 -----LATGDSKESSTANGCFTTQAGDIAHYI--DTST 967
QY 676 LKNIGLPENGELVYVFRQTQSSGSLMAAQEVHLKOGFEFFVYKPEVEGPI-----PD 728
Db 968 SSNTGTLTSGQCPVY-----WCAPSITTSGGVYTYGVIANGVIFTLVLRAD 1015
QY 729 GSVYRATLHAFVN 740
Db 1016 GSISFTLNGAVD 1027

RESULT 15

AD1374

Internalin proteins, probable peptidoglycan bound protein (LXRTG motif) homolog lmo23

C:Species: Listeria monocytogenes

C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001

C:Accession: AD1374

R:Glaser, A.P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;

Ok, C.; Schlutener, T.; Simoes, N.; Tillet, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AD1374

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-940 <GLA>

A:Cross-references: GB:NC_003210; PIDN:CAD00474.1; PID:g16411884; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:
A:Gene: lmo2396

Query Match 3.28; Score 142.5; DB 2; Length 940;
Best Local Similarity 17.58; Pred. No. 1;
Matches 139; Conservative 110; Mismatches 254; Indels 291; Gaps 34;

QY 239 NMPGNPDLMLTOSQVDAVATLMDVASASMSFEYENGSGTYSYVVGALNNRERYKSL 298
Db 45 NIICNPFLDEITENM-----ATTVADLSNMMSGAPGYGLKIDLTGLDKA- 92
QY 299 QLAHVALYTSQEH---DMIR-----GELASGRPYVYAGNNS 333
Db 93 -VNNTKLYFSNQSQIKMLDKIKLNLKIIAATVTTGLNINISALGEMALEVELEGD--- 148
QY 334 IGHAFVCDGYASDGT-----PHFN--WGMG-----VSNGFYK----- 364

LT 16
76
Surface protein (LPTXG motif) homolog lln0352 [imported] - *Listeria innocua* (strain
C:Species: *Listeria innocua*
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: A11476
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fehl, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krett, J.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tillet, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of *Listeria* species:
A:Reference number: AB1077, MUID:21537279, PMID:11679669
A:Accession: A11476
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1349 <GLA>
A:Cross-references: GB:LM592022, PIDD:GAC95585.1; PIDD:g16412781, GSPDB:GN00178
A:Experimental source: strain Cl1p11262
C:Genetics:
Gene: lln0352

A;Gene: 11n0352

```

QY      47 TAVSDKSIDVYNOGDAE-----RGITSOEGSPAY---FYVANRGNNEGVALVAADD   97
Db      370 TETWNPATVINYVANYANDYKLTSTFKDAQOGELKAPVVDTRDKDHYI-----ODNYTTTAA--   422
QY      98 RIPTILASPIGRSDMDSMPDNLRMWLQIYDOIELGLISKAQAONEILTREGVPAREVHA   157
Db      423 ---TIIPGSILV-----STPAN-----OQGTFAGANVTYNNV-----         450
QY      158 LMDNGHPANDPMERNOCYPMNPN---KEPLLPGNHAYTCGVATAAQIRYH---SWPLQ   211
Db      451 -----YKKDDYLLSTYYKDTNGCELKAPVADATTYHYQOTYTTLTAAFVGTYLLVNTPTN   504
QY      212 GEGSFEDYHAGSIYGNM-----SOTPEAMTDWIMPPENPBLDNILOSQVADYATLMR   262
Db      505 ATGFTEG---SSNITVNYTYANGYQLSTFEDKOQKTIAF---DVDPAKYTHVNDPYTTTAK   560
QY      263 DVSASVSMSEFYENGSGTYSYVVYGALNRNFRYKRSLQLAHRBALYTSOEWHDMIRGEIASG   322
Db      561 TIPTYLLVTPPANNOGFG-----JDIIVDYRYKAEEDYTLISTYKDAG   605
QY      323 R---PY-----YYAGNOSI-----GHAFVCIDGYASDGTFFHNMGWGVSNGEYKLT   367
Db      606 KELKOPVDSKRTKHIDONTSTAALIPGYLVATPANETGETFTHS---DVTVNYVVKLND   662
QY      368 LSPTSLGIGEGI-----GFTIYOELITGIIEAKTPA-EAGTDALPILA   410
Db      663 LKLTISTYKNAOGELKAPVVDSTKYTHIIDONYATTAAVIPGYLVATPANOGST----FG   717
QY      411 LMDIEAEYKSEGLNVGSIYNTGEBOSNLDLGYRLNKADGEYIEVKTSSINISMWGYGE   470
Db      718 STDIOYNYYQA---VAYKLSTYKQOGNDLA-LPKYD-----       752
QY      471 HRESFLANOLOSGINTITLLYRRKGITBOMEVRHAGCGYNSIKVNTTDBNNVYTVD   530
Db      753 -----TKTY-----HIIDGYTSS---DIAIGTYLVAAP   778
QY      531 NNE---GRLSIYPNSEFVADLNSEYHEHTTVOFNDSPODEIRTPVAPA-----LST   577
Db      779 TNQTGFEGASDVTVN-YYYKANDY---ILLSTYKNAOGKELTPIYIDOKKHINDTYTTT   834
QY      578 GATADVLSLGWMAAEVPG-----GSSNYPV-VMSKDYVLLTBSGDYTLMYRFESINNOK   629
Db      835 GATIP-----GYTLVAAPANOSGTFGANAVTVNYKYAD-----DYTLSTYKIDANGK   882
QY      630 D-----EMKKISVSXKPTLETETHPLFEVGHNOJSTY-TLDAHNRVLPE-DFT   675
Db      883 ELKAPVVDISKTYHTKDNSTSAATIPGYT-LVAAPANOGTSTFNTSNVTYVYVYKANXYT   940
QY      676 I---KNL-----GLPFNGELVYVVFROTOSSGSLMAOCTY   708
Db      941 LTSFEKNVGOELKPAIVKGFIIKDGVAITSGVITIPGTYLVA--TPSKRKGTFGASANTV   998
QY      709 HIKOGETPVYK   719
Db      999 N-----YVYK   1003

```

RESULT 17
J00420
beta-1,3-glucanase A1 precursor - *Bacillus circulans*
C:Species: *Bacillus circulans*
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 15-Oct-1999
C:Accession: J00420
R:Yahata, N.; Watanabe, T.; Nakamura, Y.; Yamamoto, Y.; Kamimura, S.; Tanaka, H.
Gene 86, 113-117, 1990
A:Title: Structure of the gene encoding beta-1,3-glucanase A1 of *Bacillus circulans* W
A:Reference number: J00420; MUID:90185240; PMID:2311931
A:Accession: J00420
A:Molecule type: DNA
A:Residues: 1-682 <YAH>
A:Cross-references: GB:MA4503; NID:g142972; PIDN:AAA2474.1; PID:g142973

A:Experimental source: strain WL-12
C:Comment: This enzyme, together with chitinase, is crucial for hydrolyzing yeast and fly
C:Genetics:
A:Gene: glcA
F:1-38/Domain: signal sequence #status predicted <SIG>
F:39-682/Product: beta-1,3-glucanase A1 #status predicted <MAT>

Query Match 3.2%; Score 142; DB 2; Length 682;
Best Local Similarity 19.8%; Pred. No. 0.66; Mismatches 197; Indels 184; Gaps 32;
Matches 114; Conservative 82;

QY 308 SQEHMDIRKELASGRPVY-AGNNOGIGHAVCDGASDGFHNNG-WG-GVSNFYK 364
DB 78 SAFMND-----VSDGVNKKVGNMWDIDAGGYTNQNGHSDGFGNW- 125
QY 365 LTLSPSLGIGGIGIF-----TIYOEI-----ITGIEPAKTPAEAGTDLPIALDIE 415
DB 126 FTLSATTEIQLYSFANGVLEOLVFONINKTTITANNPTOGP-----QIT 171
416 AEYSESGLVNGYSIVNTEGOSNLDGLRLKADGEVIEKTSIN-ISMVYGEHPES 474
172 ASFTGGAGFT--YTFNNDKAVT-----YEAADDLKYYKVPVNSSHIDIDNNAAS 221
QY 475 FSLAPNLSGINTTLLYRRTGEQEPVRAAGGYNSIKVTTDPNNVYTVDNNEG 534
DB 222 GWYDHNFGQ-----FTDGGGYWFNTES-----INVLESKTSANLVYITFNEP 269
QY 535 KLSIVRSFYADLNSYHSHITTVQFNSDPEITPVALSTGATADVISLGMVAEV 594
DB 270 ----TRNSYV--ITPYGFTTADANG-----SIDIPIPKI 299
QY 595 PGSSNPVYVMSKDVLTSEADYTLWRFSSINNOCKMKKIGSVKTPREYTHLPREV 654
DB 300 DCGA--PI--AKEL-----GNEY--YQINNGQ--W-----VDLSNDSOSKRAIYANG 339
QY 655 HNOGST-----YTLDMAN--RYLPDTKLNGLPENG-----LVVYRQTOSS 697
DB 340 YNNNSDANQGYMADYIYGLMFQPIQENMOIRIGYPLNGAGNIGNVYVYTGPNNA 399
QY 698 SGSLMAQOEYVHK-----QGETFYKPYVESEPIPD-----GSY----- 731
DB 400 PRDVSQOEDISIGTPPDPAIAGMNLIMODEFNGLTLDTSKNMETEYLLNNDPATWGW 459
QY 732 RATLHAFVNGOQOYLTK-GKRNVTYKLVNGTAVEAISSEIRVFPAPADYVEISAPCI 790
DB 460 NAELOHTNSTQNTYVDDKLN-----IKAMNDSKSFPODPNRYAQS----- 502
791 PETSILLFDLSGRIVKNSLSAGHGMV-SRLPNG 826
503 -----SKRINTKDKLTKYGRVDFRAKLPTG 528

RESULT 18

T20492
hypothetical protein F02C12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T20492
R:Colles, L.
Submitted to the EMBL Data Library, September 1995
A:Reference number: 219282
A:Accession: T20492
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1827 <WIL>
A:Cross-references: EMBL:Z54265; PIDN:CAA91022.1; GSPDB:GN00028; CESP:F02C12.1
C:Genetics:
A:Gene: CESP:F02C12.1
A:Map position: X
A:Initons: 34/1; 90/1; 113/2; 182/3; 253/3; 301/3; 385/3; 476/3; 521/3; 559/3; 598/1; 624
C:Superfamily: Caenorhabditis elegans hypothetical protein F02C12.1

Query Match 3.2%; Score 141.5; DB 2; Length 827;
Best Local Similarity 20.5%; Pred. No. 0.96;
Matches 127; Conservative 85; Mismatches 239; Indels 169; Gaps 30;

QY 334 IGHAFCVCDGASDGFHNNMGVSNNGFYKLLTSLGIGGIGFTYOEITGIE 393
DB 155 IGAIVTQKYIGIGLARLDW-----FQIDKDEEQMKNLGRKIGFAVYQ-----Y 199
QY 394 PAKTPAAGDIALP-----ILALDIAE-----YKSESLNV-----GSITNMGEEQ 437
DB 200 PTNLPOEERNSLMDQLOELMKTMLKAEKANNNSKSKKIDLFVEPEFT-----EQ 254
QY 438 SN-----LDGYRLNK-ADGEVIEKTSINISMVYGEHPESFSLAPNLSGIN 487
DB 255 EDHILKRLDELFDINRQKKOKEESNIVEKFPVFPADGSHNLIPLICLRIS--D 311
QY 488 TITLLYRRTGT--BOMEVPRRAAGGYNSIKV--NTDPNNVYTVDNNEKSLIYVNS 542
DB 312 TKIMYESTGAYDWE-----DFLNKRLPCNLIYPDGLYEV--QKIKLKENH 361
QY 543 FVADLNSYHSHITTVQFNSDPEITPVALSTGATADVISLGMVAEVPDGSNYP 602
DB 362 FVFSWNEEYVLTGAMTFLEVSFQISDRSANGIGVAATIGSVIGLFTPLAVVAATTLTYT 421
QY 603 VVMSKDVLTSEGDYTLWRFSSINNOCKD-----EMKKI-----GSVSYKT 642
DB 422 AI-GTGYSMARSGHYHVDRIHNNENVPILSRNPDMVLAAVSLASFGAIGSATL-- 478
QY 643 PTEYTHLPFEYHNOJSTYTLDMANRYLPDTKLNGLP-----FNGELVYVFR 692
DB 479 -TVMTQGLEV--SQALEYVYVNA--IFANFVSGIALTFASGNIYFEKVANGERPIPLE 532
QY 693 QTOSSSSSLMAQOYTHIKOGEFYKPYVNGPIPDGSRATLAFVNGQ----- 742
DB 533 LFQSTSLFPTNAVNLQTAEKLIQOTQO-KIND--YRQTLTAEDKQFQDKINEM 589
QY 743 ----QOYLTKGRNY--TVKIVNGTA-----VEAIESEEI--RVFPAPADYVEISA-- 787
DB 590 KNTGCOGNSKMKANLEITIDIHKAMLADKVEVHEHVDYPRKSKPSPSSVYKMAQK 649
QY 788 -----PCIPQETSILL--FDLSGRIVKNSLSAGHGRMD 819
DB 650 LTVSDKRIVEEHLQALDARGSKLPASQKITTYLVNGMDIAKV--KALLSGAGGAL 707
QY 820 VSRLPNGAYILKVDGYTKI 839
DB 708 TERL-----IDGKTKV 719

RESULT 19

A90428
serine protease, subtilase family, probable [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: A90428
R:She, Q.; Singh, R.K.; Conaloni, F.; Zivanovic, Y.; Allard, G.; Ayaaz, M.J.; Ch
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi Ngoc, H.P.; Redder
arrett, R.A.; Ragan, M.A.; Sengen, C.W.; Van der Oost, J.
Submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: A90428
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1308 <KUR>
A:Cross-references: GB:AE006641; NID:913815855; PIDN:AAK42680.1; GSPDB:GN00155
C:Genetics:
A:Gene: SSO2551
Query Match 3.2%; Score 141.5; DB 2; Length 1308;
Best Local Similarity 18.5%; Pred. No. 2;
Matches 189; Conservative 137; Mismatches 341; Indels 355; Gaps 53;

```

OY 56 DIVYRQDAERGITSQEBGSPAYFVYVARGNNEGALVAADDRIPT----- 101
DB 362 DYFFELGSAE-GIT-----FLASSGD-----LGAVNDLPYVGSANYPASSPEV 404
OY 102 -----ILASPGRFDMDSMPDLRMWLOIYDDELILISGAQOLNEET 145
DB 405 TAVGQTSIPANTSGYSTYSTNSGTFCAEI-----AM-SYNPLRYEIVGGVSS----- 452
OY 146 LRTGCVPAEVALMDNGHFANDPARNMGCPYMNKEPLDN-----GHNATYGCATAAQAQI 202
DB 453 -----GGGSQLFPAPMYORVYTHSNRAIPDVADAADNPYTGFTIYALGE 498
OY 203 MRHSMPLQEGSGSDIYAGSLVGN-----MSGTEGEMVDWIMNP-----GNDDLNLJDS 252
DB 499 V-----VIGGTSLSAPLMDGIADIDGIIHGLGLVGNPLIYETIYON 539
OY 253 QVDAYATLMDVSAVSMSFEYEN-----GSGTIVYVYVGAALRNFRYKSLQALHVAL 305
DB 540 -TTLVHOAFHQISLGYNGYVYVANSNLYVTGLASPMAGMLGVIRK-SLSKSLAISVSTF 597
OY 306 YTS--OEHMIMINGELASGRPYVYACNNOSI-----GNAFYVDGAS 345
DB 598 ETGVEQPM-----YFGSTPTIAAYITYPNNITVSGSGFNATYVSEGYLA 643
OY 346 DGFHFNMWGVSNCFYKLTLLSPSL-----GIG-----GEGI----- 380
DB 644 TVPLSFN---GSYVWGNITITPNNPMLMEIYVNGSSDQFTGVGVVEVDGESINIVSPI 700
OY 381 ---GFTIYOEIITGE-----PAKTPA-----EA 401
DB 701 PYPSPPIPYNSPFIEAMITYPNCTPYVQNSVATAYLVNDCKLLASIPILTMAPGLYEG 760
OY 402 GTDALPILA---LNDIAEYKSESGLVGYSIYNTGEBOSNDLGYRLKADGEVY--- 454
DB 761 SYALLPPLPOGYLLIYVNDYGSASFY-VYGEYNEFALITLPINDGPPA-ASPGONITIT 818
OY 455 -EVKTSI-----NISWYIGEHP-----SFSLANQLSOGINTITLLYRTGTEOME 502
DB 819 DEVLPELTGLFTSWATAIYINQNLIDQVKLTPAPDEIQGVYLLFLYIANFTIPD 878
OY 503 PVRIAGGYVNSIKVNTDPNNVYVTVDNNEKLSIYNSFVADLN-----SYEH 552
DB 879 ---ASPGFYNNVIOISISNTSTGLYKAD-----FTSPFYVSPANLTLNVKNNVYEG 927
OY 553 STTVQNSDSPDEIRPAPFA-----LSTGATADVI-----SLG-WY-M 591
DB 928 ELKIFANITYPN--GTPVAKGMEFATITLPTSLNTEQLIIFGEAGIPLOYNSTLGEWGI 985
OY 592 AEVPG-----GSSNYPVY--WSKDVLTLSGDTLLM-----YRSINNO 628
DB 986 YSIPSIFYGSIFFQSGSVYSLAGPMNVYSGVSMNGTNLSTPSSFNFNVPYTF-INNI 1044
OY 629 KDEMKIGSVSVTPTEYTHPLFEVGHNOTSTYLLDMAHNRVLPD-FTLKNLGLPENGEL 687
DB 1045 VYSSKSLDSPLSKINSTTYMLSNVKSNNITINGCNVILSNVIANITYVAKNSNMTITSST 1104
OY 688 V---VVFROTOSSSGLMAOETVHIKGEFEFVYKPVYEGPIPGSGYRATLHAFVNGQO 744
DB 1105 INQVLVLDNSVSIISK--IGGDNTAVVANDSN--TIVSSYIDSKY-----AFLDPNSV 1156
OY 745 LYLKGRMYTVYVNGTVAEAIESSEIIRFPNPARDYVEISAPCIPOETSIILPLDSK 804
DB 1157 ISLSC-----VNMATNVISLSI-----PAPRIYLLSTJNVITSKES--ITVNTGE 1200
OY 805 I-----VAKNSLSAGHG-----RMDVSRIPKCAVILKY---DG--YTTKINI 841
DB 1201 YLRLLGVSMNKPVGYSVSISSPSISLSTIPNASOLDQYTFYVISIDGLPYNLTEFL 1260
OY 842 VH 843
DB 1261 LN 1262

```

RESULT 20
 S76851
 hypothetical protein - *Synechocystis* sp. (strain PCC 6803)
 C:Species: *Synechocystis* sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence, revision 25-Apr-1997 #text, change 08-Oct-1999
 C:Accession: S76851
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasu, A.
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp.
 A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S76851
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-1771 <KAN>
 A:Cross-references: EMBL:D90917; GB:AB001339; NID:g1653836; PIDN:BAAL8763.1; PID:d101
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 3.2%; Score 141.5; DB 2; Length 1771;
 Best Local Similarity 19.9%; Pred. No. 3.1;
 Matches 103; Conservative 95; Mismatches 262; Indels 299; Gaps 41;

```

OY 95 ADDRIPITLAYSPIGRFDMDSMPDLRMWL-----QTYDEIGLILISGAQOLNEETIRTE 149
DB 347 ADIRISTIGLYA-----DNLALISTPESQIYIQRIGV----- 379
OY 150 GVPVAEVHALMDNGHFANDPARNMGCPYMNKEPLDPGNHAYGVATAAQAQIMRYHWP 209
DB 380 -----DFEINP-----PLVAGQSPNLGSLVSLGQ----- 405
OY 210 LQSGSPDYHAGSLVGNMSTFGEMYDWIMNPENPLDLNLTOSQVDAVATLM---RDVS 265
DB 406 ---SEEV-----DFPDWLQ--GSKGGEISAVRKIDGAVTLIPYIGDLA 444
OY 266 ASVSM-----SEFNGSGTYSVYVYVGAALRNFRYKSLQALHVALYISOEW--HDM 314
DB 445 AKTSILGFPAEBSRSLNDALIKGTGV--LKTWTYRTPTDY-----ITNGEFSIDAR 498
OY 315 IRELASGR---PYV-----YAGNOSIGHAFVVDGASDGFH---FNMWGCY 358
DB 499 LNFSLRSGARLNPPIPLGLQLAGANAMIKTA---APHPTAQPPEYGFYAGVVDV 555
OY 359 SNGFYKLTLLSPSLGIGEGIGFTIYOEITGIEPAPTAEGATDALPILAKDIAEY 418
DB 556 NTLFFFDST-----GFRVGFQKGVYIKGEDIKGFGIATVYDLDDEGLL-----A 601
OY 419 KSESGLVNGYSIYNTGEBOSNDLGYRLKADGEVIEKTSINISY----- 465
DB 602 TSESAL--AAVEBQNNAGDQLANGSFVNOGR-----EWASVISIRLRRGRETPAAMF 652
OY 466 -----YCGGEH-----PE-----SFLSPLNOL 482
DB 653 NLRSPSGVYVYRXYDLOGTRVKLMNEESDEFGTSLFIPREACTWYVEOYRSDDIOPI 712
OY 483 SGGINTITLL-----YRTGTEOMEVPRHAGG--YNSIKTV--TDPNN---VVVT 528
DB 713 KAVONSIAPIVAEIKQFARDNRQMWDLIYVTDPPRAENIKVLYADTGKGGDEFLIT 772
OY 529 VDNH-----EGKSIYVNS--FVADLNSYESTITVVOFNSDSPDEINTPAPFALS 576
DB 773 IDHDPKSSGSGYRRYRMDLSNVPQSVHYLYASABEGHLPDRKFTFD-----LSVATS 824
OY 577 TGATADVVISLGVMAEVPQSGSNYVYVMSKYVLLISEGDTYMTWRFSGINNOKDEMKIG 636
DB 825 TGSATDSNLPITGIEL--GCGNDLPPV--KOL-----SAN-----WIGDG 860
OY 637 SVSUKPTTEYTHPLFEVGHNOTSTYLLDMAHNRVLPDFTLKNLGLPENGELVYVVFROTQS 696
DB 861 KRSVTWSDATAPQYVYVWMSNPNAGVLD--NPDDPPS-----GI 900
OY 697 SSGSLMAOETVHIKGEFEFVYKPVYEGPIPGSGYRATLHAFVNGQOOLYLKGRMYTVK 756

```

Db 901 TSGLV-----TIEEEDGRFYHRLIVDGLTEGELYRFOIQS-VDDNDRLGAVGKTSILAV- 953
QY 757 IVNGTAVEAIESSEIRVFPNPARDYVEISAPCIPQETS 795
Db 954 AGNYDAIALGESDE-----WEYVAVLGETYTYREVS 984

Search completed: May 29, 2003, 13:27:24
Job time : 35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 13:23:06 ; Search time 15 Seconds
(without alignments)
2330.970 Million cell updates/sec

Title: US-10-030-330-1

Sequence: 1 MKKSFLLAVLMFGIAMQGH.....PNGAVILKVDGTTKRNIVH 843

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 112892

Maximum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1198	27.0	868	1 PRTT_PORGI	P43158 porphyromon
2	494	11.1	398	1 SPEB_STRPY	P00788 streptococ
3	179	4.0	939	1 EAE_ECO27	P19809 escherichia
4	154	3.5	1325	1 YDEK_ECOLI	P32051 escherichia
5	146	3.3	1409	1 HAP1_HAEIN	P44566 haemophilus
6	144.5	3.3	1256	1 MRP_STRSU	P32653 streptococ
7	142.5	3.2	1321	1 AB11_MOUSE	O9qy30 mus musculi
8	142	3.2	682	1 E13B_BACCI	P23903 bacillus ci
9	139.5	3.1	814	1 KEX2_YEAST	P13134 saccharomyc
10	139.5	3.1	1398	1 PLS_PYREU	P08686 clostridium
11	138	3.1	1398	1 PLS_PYREU	P72186 pyrococcus
12	136.5	3.1	3354	1 CADN_HUMAN	O9h251 homo sapien
13	136	3.1	426	1 GUN2_ERWCH	P07103 erwinia chr
14	136	3.1	490	1 TAY3_TREDE	P18164 treponema d
15	135	3.0	987	1 YD94_METJA	O58789 methanococ
16	134.5	3.0	976	1 AMY_BUTFI	P30269 butyrivibri
17	133	3.0	837	1 HFC1_HAEIN	P33397 haemophilus
18	133	3.0	934	1 EAE_ECO57	P43261 escherichia
19	132.5	3.0	802	1 CSD2_ECOLI	P53513 escherichia
20	132.5	3.0	1103	1 VQ37_BPARI	O9g005 bacterioph
21	130.5	2.9	1286	1 AIDA_ECOLI	O03105 escherichia
22	130	2.9	762	1 SLAP_ACEKI	P22258 acetogenium
23	128	2.9	914	1 GUN2_CLOSR	P50990 clostridium
24	127.5	2.9	3317	1 CADN_PAT	P58365 ratius norv
25	125	2.8	837	1 HFC3_HAEIN	P45998 haemophilus
26	125	2.8	878	1 FIMD_ECOLI	P30130 escherichia
27	124.5	2.8	820	1 CHIA_ALTSO	P32883 alteromonas
28	124.5	2.8	865	1 HTRE_ECOLI	P33129 escherichia
29	124	2.8	1541	1 IGAL_HAEIN	P42782 haemophilus
30	123.5	2.8	852	1 CSG_HALHA	P08198 halobacteri
31	123.5	2.8	862	1 SLA2_BACAN	P94217 bacillus an
32	123	2.8	676	1 HMUR_YERPE	O56989 yerusalemia
33	123	2.8	702	1 FOXA_SALTY	O56145 salmonella

34	123	2.8	935	1 EAE_ECOLI	O31000 escherichia
35	123	2.8	1259	1 YTFN_ECOLI	P39321 escherichia
36	122.5	2.8	1218	1 MGPC_MYCPN	O50341 mycoplasma
37	122.5	2.8	1609	1 FTG2_YEAST	P25653 saccharomyc
38	122	2.7	760	1 YBIL_ECOLI	P75780 streptococ
39	122	2.7	936	1 EAE_CITER	O07591 citrobacter
40	121.5	2.7	1628	1 NAGH_CLOPE	P26831 clostridium
41	121	2.7	837	1 HFC2_HAEIN	P45997 haemophilus
42	121	2.7	1167	1 CLAA_BACTU	P56956 bacillus th
43	120.5	2.7	819	1 CSD1_ECOLI	P53512 escherichia
44	120.5	2.7	972	1 CTAL_BACCI	P94286 bacillus ci
45	120	2.7	1953	1 BIGA_SALTY	P25927 salmonella

ALIGNMENTS

RESULT 1	ID	PRTT_PORGI	STANDARD	PRT	868 AA.
AC	P43158	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	01-NOV-1995 (Rel. 32, Last annotation update)				
DE	Thiol protease/hemagglutinin prtt precursor (EC 3.4.22.-).				
GN	PRTT.				
OS	Porphyromonas gingivalis (Bacteroides gingivalis).				
OC	Bacteria; Bacteroidetes; Bacteroides; Bacteroidales;				
OC	Porphyromonadaceae; Porphyromonas.				
OX	NCBI_TaxID=837;				
RN	SEQUENCE FROM N.A.				
RC	STRAIN=ATCC 53977;				
RX	MEDLINE=93114862; Pubmed=8093357;				
RA	Otago J. I., Kuramitsu H.K.;				
RT	"Isolation and characterization of the Porphyromonas gingivalis prtt gene, coding for protease activity."				
RL	Infect. Immun. 61:117-123(1993).				
RN	REVISIONS, SEQUENCE FROM N.A.				
RC	STRAIN=ATCC 53977;				
RX	MEDLINE=95105001; Pubmed=7806362;				
RA	Madden T.E., Clark V.L., Kuramitsu H.K.;				
RT	Revised sequence of the Porphyromonas gingivalis prtt cysteine protease/hemagglutinin gene: homology with streptococcal pyrogenic exotoxin B/streptococcal proteinase.;				
RL	Infect. Immun. 63:238-247(1995).				
CC	-1- FUNCTION: APPEARS TO BE SPECIFIC FOR ARGININE-CONTAINING PEPTIDE BONDS. POSSESSES HEMAGGLUTININ ACTIVITY.				
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C10.				
CC	-1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-17 IS THE INITIATOR.				
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	EMBL: M83096; -; NOT_ANNOTATED_CDS.				
DR	MEROPS: C10.002; -				
DR	InterPro: IPR000200; Peptidase_C10.				
DR	InterPro: IPR001230; Prenyl_site.				
DR	Pfam: PF01640; Peptidase_C10; 1.				
KW	PRINTS: PR00797; STREPTOPAIN.				
KW	HydroLase: Thiol protease; Signal; Hemagglutinin.				
FT	STGNL	1	27		POTENTIAL.
FT	PROPEP	28	?		THIOL PROTEASE/HEMAGGLUTININ PRTT.
FT	CHAIN	184	868		BY SIMILARITY.
FT	ACT_SITE	184	868		BY SIMILARITY.
FT	ACT_SITE	327	327		BY SIMILARITY.
SO	SEQUENCE	868 AA;	96444 MW;	45436EFP3279323 CRC64;	

RT cyanogen bromide peptides: complete covalent structure of the
RT polypeptide chain.";
RT J. Biol. Chem. 251:1955-1959(1976).
RN [9]
RP FUNCTION.
RC STRAIN-N2131 / Serotype M49,T14;
RX MEDLINE=99081733; PubMed=5864206;
RA Kuo C.-F., Wu J.-J., Tsai P.-J., Kao F.-J., Lei H.-Y., Lin M.T.,
RA Lin Y.-S.;
RT "Streptococcal pyrogenic exotoxin B induces apoptosis and reduces
RT phagocytic activity in U937 cells.";
RL Infect. Immun. 67:126-130(1999).
RN [10]
RP FUNCTION.
RC STRAIN-A-20 / Serotype M1,T1;
RX MEDLINE=99368817; PubMed=10456871;
RA Tsai P.-J., Lin Y.-S., Kuo C.-F., Lei H.-Y., Wu J.-J.;
RT "Group A Streptococcus induces apoptosis in human epithelial cells.";
RN Infect. Immun. 67:4334-4339(1999).
CC -1- FUNCTION: IMPORTANT STREPTOCOCCAL VIRULENCE FACTOR WHICH CLEAVES
CC HUMAN FIBRONECTIN AND DEGRADES VITRONECTIN. ALSO CLEAVES HUMAN
CC IL1B PRECURSOR TO FORM BIOLOGICALLY ACTIVE IL1B. CAN INDUCE
CC APOPTOSIS IN HUMAN MONOCYTES AND EPITHELIAL CELLS IN VITRO, AND
CC REDUCES PHAGOCYTIC ACTIVITY IN MONOCYTIC CELLS. THUS, MAY PLAY A
CC ROLE IN BACTERIAL COLONIZATION, INVASION, AND INHIBITION OF WOUND
CC HEALING.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage with hydrophobic
CC residues at P2, P1 and P1'.CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DISEASE: THE STREPTOCOCCAL PYROGENIC TOXINS A, B, AND C ARE
CC THE CAUSATIVE AGENTS OF THE SYMPTOMS ASSOCIATED WITH SCARLET
CC FEVER. HAVE BEEN ASSOCIATED WITH STREPTOCOCCAL TOXIC SHOCK-LIKE
CC DISEASE AND MAY PLAY A ROLE IN THE EARLY EVENTS OF RHEUMATIC
CC FEVER.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C10.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial/
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M86905; AAA26978.1; -
DR EMBL: L26125; AAA26979.1; -
DR EMBL: L26126; AAA26982.1; -
DR EMBL: L26127; AAA26993.1; -
DR EMBL: L26128; AAA26994.1; -
DR EMBL: L26130; AAA26996.1; -
DR EMBL: L26131; AAA26997.1; -
DR EMBL: L26132; AAA26998.1; -
DR EMBL: L26133; AAA26999.1; -
DR EMBL: L26134; AAA27000.1; -
DR EMBL: L26135; AAA27001.1; -
DR EMBL: L26136; AAA27002.1; -
DR EMBL: L26137; AAA27003.1; -
DR EMBL: L26138; AAA27004.1; -
DR EMBL: L26139; AAA27005.1; -
DR EMBL: L26140; AAA27006.1; -
DR EMBL: L26141; AAA27007.1; -
DR EMBL: L26142; AAA27008.1; -
DR EMBL: L26143; AAA27009.1; -
DR EMBL: L26144; AAA27010.1; -
DR EMBL: L26145; AAA27011.1; -
DR EMBL: L26146; AAA27012.1; -
DR EMBL: L26147; AAA27013.1; -
DR EMBL: L26148; AAA27014.1; -
DR EMBL: L26149; AAA27015.1; -
DR EMBL: L26150; AAA27016.1; -
DR EMBL: L26151; AAA26980.1; -
DR EMBL: L26152; AAA26981.1; -
DR EMBL: L26153; AAA26982.1; -

DR EMBL: L26154; AAA26983.1; -
DR EMBL: L26155; AAA26984.1; -
DR EMBL: L26156; AAA26985.1; -
DR EMBL: L26157; AAA26986.1; -
DR EMBL: L26158; AAA26987.1; -
DR EMBL: L26159; AAA26988.1; -
DR EMBL: L26160; AAA26989.1; -
DR EMBL: L26161; AAA26990.1; -
DR EMBL: L26162; AAA26991.1; -
DR EMBL: AF104940; AAD17930.1; -
DR EMBL: AB030578; BAB16027.1; -
DR EMBL: AE006625; AAK34706.1; -
DR EMBL: AE010112; AAL98559.1; -
DR EMBL: L26129; AAA26995.1; -
DR PIR: A37768; A37768.
DR PIR: A00978; BRSOP.
DR MEMOPS: C10.001; -
DR InterPro: IPR000200; Peptidase_C10.
DR Pfam: PF01640; Peptidase_C10; 1.
DR PRINTS: PRO0797; STREPTOPAIN.
KW Hydrolyase; Thiol protease; zymogen; Toxin; Signal; Complete proteome.
FT SIGNAL 1 27
FT PROPEP 28 145
FT CHAIN 146 398
FT ACT_SITE 192 192
FT ACT_SITE 340 340
FT VARIANT 8 8
FT VARIANT 17 17
FT VARIANT 80 80
FT VARIANT 111 111
FT VARIANT 137 137
FT VARIANT 154 154
FT VARIANT 193 193
FT VARIANT 211 211
FT VARIANT 293 293
FT VARIANT 305 305
FT VARIANT 308 308
FT VARIANT 317 317
FT VARIANT 384 384
FT VARIANT 394 394
FT CONFLICT 84 85
FT CONFLICT 169 169
FT CONFLICT 187 191
Query Match 11.1%; Score 494; DB 1; Length 398;
Best Local Similarity 30.6%; Pred. No. 5e-23;
Matches 129; Conservative 73; Mismatches 155; Indels 64; Gaps 13;
OY 6 LIAIYMGIGIMOG-----HSAPYTKRALSLARLARLOYSLRGQAVSPKISIDVY 59
DB 10 LLSIALGCFYLANVFPDQNFARKEKAKBSAIFFOKSAIKAGASAEI-ITLDKVN 68
OY 60 RQGAERGITSOEGSPAYFYVANGNEGVALVAADRIPITILAYSPIGRFDMSPDN 119
DB 69 LGGEL-----SGSNMYVYNISG---GFYIVSGDKRSPILIGYSTGSPDANG-KEN 116
OY 120 LRMLQIT-----DQELGLISGAQUNELRLREGYPAEYHALMDNGHRANDPMRNO 173
DB 117 IASFESYVEQIKENKLDITTYAGTAEIKOPV-----VKSILD-----SKGIHYNQ 162
OY 174 GYPNNKEPLLPN-----GNHAYGCVATAAQTMRYSNPLQEGSGFDY----- 218
DB 163 GNPVLLLPVLEKAPGQSGSFVGGHAATGCAVATATQIMKHNTYKRLKDTYTLSSNN 222
OY 219 ---HAGSLVGNWSGTFGEWTDWIN-MGNDPLDNLTSQVDAYATLIRDVASVSNSEY 273
DB 223 PYFNHPKML---FAAISTROYNMNIIPLTYSGRSESNVQKM--AISELMADVGISVDMDYG 277

DT 01-NOV-1997 (Rel. 35, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Hypothetical lipoprotein ydek precursor (ORF1).
 GN YDEK OR ORF1 OR B1510.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 NCBI_Taxid=562;
 RN NCBI_Taxid=562;
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 [2]
 [3]
 SEQUENCE FROM N.A.
 MEDLINE-97251357; PubMed-9097039;
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakada S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
 RA Sempel G., Seki Y., Sivasubramanian S., Tagami H., Takeda J.,
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiiuchi T.;
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 RN [3]
 RP SEQUENCE OF 595-1325 FROM N.A.
 RX MEDLINE-94100243; PubMed-8274505;
 RA Carverly P.J., Tims M.W., Lithgow T., Hoef P.B., Hoogenraad N.J.;
 RT "An Escherichia coli gene showing a potential ancestral relationship
 to the genes for the mitochondrial import site proteins ISP42 and
 MOM38.";
 RL Biochim. Biophys. Acta 1153:345-347(1993).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 (potential).
 CC -1- SIMILARITY: TO E. COLI YDEK.
 CC -1- SIMILARITY: SOME, TO FUNGAL MITOCHONDRIAL IMPORT SITE PROTEINS
 ISP42 AND MOM38.
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
 FRAMESHIFT IN POSITION 653.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000248; AAC74583.1;
 DR EMBL: D90793; BAA15190.1; ALT_INIT.
 DR EMBL: D90794; BAA15197.1; ALT_INIT.
 DR EMBL: X73295; CAA51730.1; ALT_FRAME.
 DR PIR: S34315; S34315.
 DR ECGene: EG11780; ydek.
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Hypothetical protein; Membrane; Lipoprotein; Signal;
 KW Complete proteome.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 1325 HYPOTHETICAL LIPOPROTEIN YDEK.
 FT LIPID 19 19 N-ACYL DIGLYCERIDE (POTENTIAL).
 FT CONFLICT 884 884 N -> K (IN REF. 3).
 FT CONFLICT 1317 1317 M -> S (IN REF. 3).
 SQ SEQUENCE 1325 AA; 136514 MW; 26A3A065FA19AD7D CRC64;
 Query Match 3.5%; Score 154; DB 1; Length 1325;
 Best Local Similarity 21.7%; Pred. No. 0.14;

Matches 157; Conservative 88; Mismatches 260; Indels 220; Gaps 39;
 QY 161 NGFAPDPMRWNGCYPMNNKEPLL---PNCNHAHYTCVATAAQMRYHSPLOGEGSFD 217
 Db 295 NGYFGNGVTNINNNGLNNKXSLVGVDSHCVVNTDK----- 334
 QY 218 YHAGSLVGNMS--GTFGE--MYDMINPAGPDLDNLTQSDVADATLMRDVSAVSMSFY 273
 Db 335 -----GHNWFLGT--GEARYIYIGDAGGELNWSSEKAVD-----SGITTAGMK 377
 QY 274 ENGSGTYSV---YVVGALRNFRYKSLQLAH--RALYTSQEHMDMIRGELASGRPVY 327
 Db 378 EGTGNTITVADKNSVITNLTGNTGCGHEMNINSGLVYSSNGSSLSGCTEYGANVSIT 427
 QY 328 AGNNQSIG-HAFVCDGYASDGTFFNMG-----WGVSNGRFKLLTSPTS---- 372
 Db 438 TGMMEVKNKVVYTTIGVAGVGNINISDGVKFSVSONITFLDKASGIGTLMMATSEFDT 497
 QY 373 ---LGIGGEGI-----GFTIYE---ITGIEPAKTPAAGTDALPLAKDIEAEYKS 420
 Db 498 VGINVGFGSGIVNVSNGATLNTSTGYGFGNAGSKCIIVISTDSMLNLTSSTNOL-- 555
 QY 421 ESSLNNGYSIYNTGEOSNLDLGRUNKADEVEIYVTSI-----NISWYGEHP 472
 Db 556 ---LVQV--VLGTGE-----LNTTGGIYKARDIOLALNDSKGVYVDGQNSL 600
 QY 473 ESFSLAPNLSGGINITLLYRRT---GTEQW---EPV-----RHAQGVY 512
 Db 601 EFTNMYVG--TSGTGTLTNNGLTINVEGGEVYLVGEPVAGTLNIGAHGAADAGFT 658
 QY 513 -NSIKY-----NTDPN---NVVYTDNNEKLSIVPSFADLNSVHSF 554
 Db 659 TNAKVEFGIGCEVFPFNTNSDAGYQVMLITGDDKDKV-----HDA 704
 QY 555 ITQFNSDSPEDEIRTPAFALSTGA--TADVISLGVMAEV---PGSSNPVYWSKD 608
 Db 705 GHTVFNAGNYSKTLVNOGLTLASHTAGVYGMG--SEVVIANG-----TLD 753
 QY 609 VL--TISEGDTYLTWRES---INNOKDEKTKIGSVKPTPEYTHPLFEVGHNNSTYT 662
 Db 754 ILASTNSAGDYTLTNLAKGDLKRVOLSSSDKKEGFHAHTGEFA---GVAOLKOSTFT 809
 QY 663 LDMAHRVLPDFTLKNLGLPENGELVVFQOTSSSSLMAAETHIKGETFV---- 718
 Db 810 LENDNRAALTHAMLOS-----DSENTTSVAVGEOSIGL-----AMNGGIIITPTDI 856
 QY 719 -RPVEGPIPDGSYRATLHAFVGOQLYLKGRNRYTVKINGTAVAELESSEIRVFP 776
 Db 857 PAATLAEGLT-----SVDTLVVAGADYTWKG--RNYQ---VNGTGVLLD-----VP 898
 QY 777 NPARD 781
 Db 899 KPMND 903
 RESULT 5
 HAP1_HAEIN STANDARD; PRT; 1409 AA.
 AC P44596;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, last sequence update)
 DT 15-JUN-2002 (Rel. 41, last annotation update)
 DE Adhesion and penetration protein precursor (EC 3.4.21.-).
 GN HAP OR H10248.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
 GN Haemophilus.
 NCBI_Taxid=727;
 [1]
 SEQUENCE FROM N.A.
 RP STRAIN-Rd / KW20 / ATCC 51907;
 RX MEDLINE-95350630; PubMed-7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,

RA Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,
 RA McKenney J.D., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrman J.L., Geoghegan N.S.M.,
 RA Greth C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.";
 RL Science 269:496-512(1995).
 CC - FUNCTION: PROBABLE: PROTEASE; PROMOTES ADHERENCE AND INVASION BY
 CC DIRECTLY BINDING TO A HOST CELL STRUCTURE (BY SIMILARITY).
 CC - SUBCELLULAR LOCATION: Secreted (Potential).
 CC - DOMAIN: THE SIGNAL PEPTIDE GUIDE THE PRECURSOR TO THE PERIPLASMIC
 CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
 CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
 CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
 CC SIMILARITY).
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
 CC - CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A STOP CODON HAD TO
 CC BE SKIPPED IN POSITION 710 TO PRODUCE THIS ORF.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U32710; NOT_ANNOTATED_CDS.
 DR MEROPS: S06.006; -
 DR TIGR: H10248; -
 DR InterPro: IPR000710; IGA_S6.
 DR InterPro: IPR004899; Pectact_sup.
 DR Pfam: PF02385; IGA1; 1.
 DR Pfam: PF03212; Pectactin; 1.
 DR PRINTS: PR00921; IGASERPTASE.
 DR HydroBase: Serine protease; Transmembrane; Zymogen; Signal;
 KW Complete proteome.
 KM Complete proteome.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 ? ADHESION AND PENETRATION PROTEIN.
 FT PROPER 7 1409 HELPER PEPTIDE (POTENTIAL).
 FT ACT_SITE 250 250 BY SIMILARITY.
 SQ SEQUENCE 1409 AA; 156797 MM; 63ABC893FA84D16E CRC64;

 QY Match 3.38; Score 146; DB 1; Length 1409;
 t Local Similarity 18.88; Pred. No. 0.47;
 cnes 177; Conservative 120; Mismatches 290; Indels 356; Gaps 47;

 QY 67 GINSOEGSPAYF-----YVANNRNGEYALVAADRIPTILAYSPIGRFDMSDPDLR 121
 DB 18 GINSQAMAGHTYICIDYRRDFAENKKEFTVCAKN----- 53
 QY 122 MMLQITDOE---IGLILSGAQLNEBITLRTGVPAEVH---ALMDNGHFANDPWRMG 173
 DB 54 --IEVYNKEGQLVGTSMTKAPMIDFSVSHNGVALVGDQYIVSAHNGY-NDVDFGAE 110
 QY 174 GYPMNKEPLLPNGNNAVTCVATAAQLMR---YHSNPLOGSGSDYHAGSL----- 223
 DB 111 G-----RNP-----DQHRFT-----YQIVKRNRYAMERKHPYDDYDHPRLAKFVTEA 154
 QY 224 --VGNMSTGEWEY-WMINMPCNPDLNLTQSOVDAYATILMRVSAVSMSFYENGSGTY 280
 DB 155 EPGMGTNMDGAKYADRENTPERVIRIGSGRQ---YMRDKDEETNHSSTIVYSGA--- 206
 QY 281 SYVYVGCALRNFRYKSLQLVHRA----- 304
 DB 207 -----RYLLAGNTHQSGNGNGTVNLSGNVSPNHYGPLPTGSKSGDSSPMF 255
 QY 305 LITSQEMHMGKELASGRPVYIAGN-----QSTGHAFCVCDGASDGTFRF 351

DB 256 IYDAKKKQWMLNVAQTGHPPFGNGFOLIREMEFYNEVLAVDFSPVFORPIPINGHX 315
 QY 352 NMWQGVSN--GYKILILSPISLGIGSEIGTITQETITTEPKATPEAGTDALPIL 409
 DB 316 SF-----VSNNDGKLTLLRPSKDG-----SKANSEGVTVLFPN 351
 QY 410 ALMDIEAEY-KSESGLVN-----GYSIYNTGEQSNLDGLRNLKADGEV-----IEY 456
 DB 352 SLNQTAKENHKAAGANNITVPRMKEYKNITLIGQSGCTLIENNINNGAGGLYFEGNPFV 411
 QY 457 KTSISINISWYGYG-----EHPESFIADNQLSQ-GINT----- 489
 DB 412 KGRNNTIWMGAGVSIQGDATVEMKVNPN-----DRLSKIGITLVNKGKKNLSLS 466
 QY 490 ----TILYRTGTEQNEPVAHAGGVN---SIKVTY---DNNVYTVVNNKGLSTY 539
 DB 467 AGNKVILDDQADEAQKQAFKEVGYVSGRATVQLNSTQVDNNNTYFGE--RGRNL--- 521
 QY 540 PMSFVADLNSYEHSITTVQFNSDSDEI-----RTPAFALSTGATADVYISLGWVAEVP 595
 DB 522 ----DLNGHSLTFRIQ-NTDEGAMIVNHTQVANNITIGNES-----ITAP 564
 QY 596 GGSNN-PPVMSKDVLTLSGDTTLMYRSINNOKDEMKGISV--KTPETYPLE 652
 DB 565 SNKNINKLDYSKEIA-----YNGWF-----GETDKRNHNGRLNLYRPTEDRTLLS 613
 QY 653 VGHNGTSTYV-----LD-----MAHNVLPDF-----TLKNG 680
 DB 614 GGNLAKGDTITQTKLFESGRPTPHAYNHLDRKWSMEGIPQGEIYWDVDMNRTKAEN 673
 QY 681 LPPNGELVVFROTQSSGSLMAAETVHIKQGETFYKPVYEGPIPDGSDYRATLHAFVN 740
 DB 674 FOIKGSAAVVSRAVSSIEGN-W-----TVSNANATP-----GVVP-----N 709
 QY 741 GQOQVLKGRN--YVYKLVNGAVAISSSEIRFPNPARDYVISAAPICPQETSIL 798
 DB 710 XQNTICTRSDMTGLTTCIKTVNLITRYI-----NSIP1 742
 QY 799 FDLGKIVKMSLSLAC-HGRMDVSRIPNGAVILKYDGYTKIN 840
 DB 743 TQINGSLINLNNATVNIHG-----LAKLNGVNTLLD 773

 RESULT 6
 MRP_STRSU STANDARD; PRT; 1256 AA.
 ID MRP_STRSU
 AC p32653;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Muramidase-released protein precursor (136 kDa surface protein).
 GN MRP.
 OS Streptococcus suis.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus;
 OX NCBI_TaxID=1307;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=serotype 2 / D282;
 RX MEDLINE=92267650; PubMed=1587602;
 RA Smith H.E., Vecht U., Gielkens A.T., Smits M.A.;
 RT "Cloning and nucleotide sequence of the gene encoding the
 RT 136-kilodalton surface protein (muramidase-released protein) of
 RT Streptococcus suis type 2.";
 RL Infect. Immun. 60:2361-2367(1992).
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 CC an amide bond (Potential).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X64450; CAA45781.1; -

DR PIR: A43829; A43829.

DR InterPro: IPR001899; Gram_pos_anchor.

DR Pfam: PF00746; Gram_pos_anchor; 1.

DR TIGRfams: TIGR01167; LPXG_anchor; 1.

DR TIGRfams: TIGR01168; YSIRK_signal; 1.

DR PROSITE: PSS0847; GRAM_POS_ANCHORING; 1.

KW Cell wall; Peptidoglycan-anchor; Repeat; Signal.

FT SIGNAL 1

FT CHAIN 48 1226 POTENTIAL.

FT PROPEP 1227 1256 MURAMIDASE-RELEASED PROTEIN.

FT REPEAT 663 681 REMOVED BY SORTASE (POTENTIAL).

FT REPEAT 839 861 SMALL.

FT REPEAT 953 1006 SMALL.

FT REPEAT 1064 1084 LARGE.

FT REPEAT 1089 1142 SMALL.

FT REPEAT 1143 1195 LARGE.

FT REPEAT 1223 1227 LARGE.

FT SITE 1226 1226 LPXG SORTING SIGNAL (POTENTIAL).

MOD_RES 1226 1226 AMIDE-LINKED TO CELL WALL (POTENTIAL).

SEQUENCE 1256 AA; 135795 MW; DCF765242F14341 CRC64;

Query Match 3.3%; Score 144.5; DB 1; Length 1256;

Best Local Similarity 18.7%; Pred. No. 0.49;

Matches 214; Conservative 121; Mismatches 374; Indels 433; Gaps 60;

QY 21 SAEVTERALSLRLALROVSLR-----MGQAVSDKISIDVYRQDAERGITS--- 70

DB 83 SAVALTSTEVEKAKAVLEQVTSPLLAGLQKELAKTEDATLAKAIEDAQTKLAAKAI 142

QY 71 -----OEGSPAFYVYANR--GNN-EGYA---LVAAADRIPITLAVSPGRDMD 114

DB 143 LADSEATVEQEVAQVAAVKAANBALGNELOKTYVGLTALADTV----- 187

QY 115 SMDNLRMALQIYDOEIGLISGAOLNEBILRTGVPAYVHAL-MDNCHFANDPMRWQ 173

DB 188 -ADPTASTLKVDGE-GTLLDSTTATPMAEPNGAIALPHILRQDIAKASEPW-- 243

QY 174 GYPMNKEPLLPNGNNA---YTGCVATAAQIMRYHWPLOEGSEFHYAGSLVGNWGT 230

DB 244 -YFEESDLYSNKNMASTYK-----AEVDAYIYSLDNDSTTAVALAEVYSTRTGD 296

QY 231 FGEYDWINMG-----NPDNLNLQSOVD-AVALMRDVSASVSMSEFENGSTSVY 283

DB 297 VLEKY--TIEPGESVTFSHPTKVNANNSITVTYDLSLASANPGALKFSAN-DDVYSTI 353

QY 284 VQALR-NMRY-----KRSIQ-----HYRALYTSQEW---HDMIRGELA 320

DB 354 IVPAYQINTRYVYESGKVLATYGIOTIAGOVVTPSSVR-VFGYDVAATTKAVQSPY 412

QY 321 SGRPVYAG-----NOSIGHAVCDGYASDQTEHFNMGCV-SNGF 362

DB 413 KG--TVYLAGVQKDVQYKIREIVENDQA---VLKFFYLDPTYGVDWNRGTDTGF 466

QY 363 YKTLSPSTSLIGSGEGIGFTYQELTG-----LEPAKT----- 397

DB 467 IELLTSPITYKVG-----TIYDYNINSKITAPETDPTKVNWKESEONEGSKYRV 520

QY 398 -----PAEAGT----- 403

DB 521 IAWMGDETTKGIYKIYINTCYWTKLGNEMGWPFSDQACIKRNKGFMPAGVQNT 580

QY 404 --DALPLAK-----DIEEYSESGLVGYSITVNGE---QSNLDIGYR 445

DB 581 LRNATPATAVETTYIKESKSGDYVIEYDGTGOKIVNSVVDTPKSALETGYNTDVR 640

QY 446 ---LNKADGEV-----IEKTSINSIW----- 465

DB 641 PASLVAAADGVYTYRKEVKSASAKTTGTVAAGTTVTKYVREKAGSVNVEVDINGKVIKAP 700

QY 466 -----YGY-----GEHP-----ES 474

DB 701 VSDEKAKAPRYNVDTDLDQKLASTFEGREKYLVPADYVGVKGNLIEVGNNTAKG 760

QY 475 FSLAPNOLSGIN-TITLLRRFTGEQMEVRRHAQGYVNSIKVN-----TDP 522

DB 761 IDPTTKIEAGVNNVEYVYRAVGVVNVKDEGNVINDPETDVSADPVDAYTTDK 820

QY 523 N-NVYVVDNN-----EGKLSI-----VNSVADLNSIEHS 553

DB 821 KPNELITKDSRYVLVPSKTDGEENGKVIETITVYVYOKVANMIRE--IPVPEPDRP 878

QY 554 TIVQFNSDPSDEIRTPVAFALSGATADVISLGWMAEYPGSSNYPYWMSKDVLTLS 613

DB 879 KVPFPDPTPEDEIDP-----TTPGNGEVPNIPY---VPG---YTVYDPRDNPPLK 925

QY 614 EGDYTLMYRFSINNOKDEMKIGSV-----SVKPTETHTPLFEVGHNOTSTYTLDMA 666

DB 926 PID-----PNDPGK--GYVPTPENPGVDPIPVY-PVKKVVTNH-----VDEE 966

QY 667 HNRVLP--DTLKNLGLP---FNGELVYVFRQTOSSGSLMAOETVHIQGE----- 714

DB 967 GNPLAPQEGSTPKPKRSIPGEFTGKV-----TDEDGNTTHIKKTPTEVNGVYVYVT 1021

QY 715 ---TFVYKPVVEGPI-PDGS-YRATLHAFVNGOOLYKGRNTYKIVNGTAVEAIESS 769

DB 1022 EDGVVKEPVTDPRTSPGTPYDT-----DNKRTIFKGE-EVELYRVGDTENGKVVEG 1076

QY 770 EEI-----RVFPNARDYV-----EISAPCIPOE-----TSILFLDSKRYKNSLSA 813

DB 1077 EYVTVYVYRREYVPAKRVYTNHVEDEGNPVAPOEGTTPKPKRSIPGEFTGKTVDEGNT 1136

QY 814 GH 815

DB 1137 TH 1138

RESULT 7

ABL_MOUSE

ID ABL_MOUSE STANDARD; PRT; 1321 AA.

AC 090Y30; 090Z8;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Bile salt export pump (Atp-binding cassette, sub-family B, member 11)

DE (Sister of p-glycoprotein).

GN ABCB11 OR BSEP OR SPGP.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxId=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

RX MEDLINE=20076398; PubMed=10607905;

RA Green R.M., Hoda F., Ward K.L.;

RT "Molecular cloning and characterization of the murine bile salt export pump.";

RT Gene 241:117-123(2000).

RN [2]

RP SEQUENCE OF 463-635 FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Liver;

RA Salkar R., Suchy F.J., Ananthanarayanan M.;

RT "Molecular cloning of mouse liver bile salt export pump (bsep).";

RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: INVOLVED IN THE ATP-DEPENDENT SECRETION OF BILE SALTS INTO THE CANALICULUS OF HEPATOCYTES.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY, IF NOT EXCLUSIVELY IN THE LIVER, WHERE IT WAS FURTHER LOCALIZED TO THE CANALICULAR MICROVILLI AND TO SUBPLASMICULAR VESICLES OF THE HEPATOCYTES BY IN SITU.

CC -1- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES, EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN

CC ATP BINDING CASSETTE (ABC) DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AF133903; AAT14372.1;
 DR EMBL: AF186585; AAD56419.1;
 DR MGD: MGI:1351619; Abcd11.
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transport.
 DR InterPro: IPR001140; ABCtransport.
 DR Pfam: PF00005; ABC_tran; 2.
 DR Pfam: PF00664; ABC_membrane; 2.
 DR ProDom: PD000006; ABC_transport; 2.
 DR SMART: SM00382; AAA; 2.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KM ATP-binding: Transmembrane; Transport.
 FT DOMAIN 1 62
 FT TRANSMEM 63 83
 FT TRANSMEM 84 147
 FT TRANSMEM 148 168
 FT DOMAIN 169 215
 FT TRANSMEM 216 236
 FT TRANSMEM 237 240
 FT TRANSMEM 241 261
 FT TRANSMEM 262 319
 FT TRANSMEM 320 340
 FT TRANSMEM 341 353
 FT TRANSMEM 354 374
 FT TRANSMEM 375 755
 FT TRANSMEM 756 776
 FT TRANSMEM 777 794
 FT TRANSMEM 795 815
 FT TRANSMEM 816 869
 FT TRANSMEM 870 890
 FT TRANSMEM 891 911
 FT TRANSMEM 912 979
 FT TRANSMEM 980 1000
 FT TRANSMEM 1001 1011
 FT TRANSMEM 1012 1032
 FT TRANSMEM 1033 1321
 FT DOMAIN 1321 1467
 FT TRANSMEM 1468 1488
 FT TRANSMEM 1489 1510
 FT TRANSMEM 1511 1531
 FT TRANSMEM 1532 1552
 FT TRANSMEM 1553 1573
 FT TRANSMEM 1574 1594
 FT TRANSMEM 1595 1615
 FT TRANSMEM 1616 1636
 FT TRANSMEM 1637 1657
 FT TRANSMEM 1658 1678
 FT TRANSMEM 1679 1699
 FT TRANSMEM 1700 1720
 FT TRANSMEM 1721 1741
 FT TRANSMEM 1742 1762
 FT TRANSMEM 1763 1783
 FT TRANSMEM 1784 1804
 FT TRANSMEM 1805 1825
 FT TRANSMEM 1826 1846
 FT TRANSMEM 1847 1867
 FT TRANSMEM 1868 1888
 FT TRANSMEM 1889 1909
 FT TRANSMEM 1910 1930
 FT TRANSMEM 1931 1951
 FT TRANSMEM 1952 1972
 FT TRANSMEM 1973 1993
 FT TRANSMEM 1994 2014
 FT TRANSMEM 2015 2035
 FT TRANSMEM 2036 2056
 FT TRANSMEM 2057 2077
 FT TRANSMEM 2078 2098
 FT TRANSMEM 2099 2119
 FT TRANSMEM 2120 2140
 FT TRANSMEM 2141 2161
 FT TRANSMEM 2162 2182
 FT TRANSMEM 2183 2203
 FT TRANSMEM 2204 2224
 FT TRANSMEM 2225 2245
 FT TRANSMEM 2246 2266
 FT TRANSMEM 2267 2287
 FT TRANSMEM 2288 2308
 FT TRANSMEM 2309 2329
 FT TRANSMEM 2330 2350
 FT TRANSMEM 2351 2371
 FT TRANSMEM 2372 2392
 FT TRANSMEM 2393 2413
 FT TRANSMEM 2414 2434
 FT TRANSMEM 2435 2455
 FT TRANSMEM 2456 2476
 FT TRANSMEM 2477 2497
 FT TRANSMEM 2498 2518
 FT TRANSMEM 2519 2539
 FT TRANSMEM 2540 2560
 FT TRANSMEM 2561 2581
 FT TRANSMEM 2582 2602
 FT TRANSMEM 2603 2623
 FT TRANSMEM 2624 2644
 FT TRANSMEM 2645 2665
 FT TRANSMEM 2666 2686
 FT TRANSMEM 2687 2707
 FT TRANSMEM 2708 2728
 FT TRANSMEM 2729 2749
 FT TRANSMEM 2750 2770
 FT TRANSMEM 2771 2791
 FT TRANSMEM 2792 2812
 FT TRANSMEM 2813 2833
 FT TRANSMEM 2834 2854
 FT TRANSMEM 2855 2875
 FT TRANSMEM 2876 2896
 FT TRANSMEM 2897 2917
 FT TRANSMEM 2918 2938
 FT TRANSMEM 2939 2959
 FT TRANSMEM 2960 2980
 FT TRANSMEM 2981 3001
 FT TRANSMEM 3002 3022
 FT TRANSMEM 3023 3043
 FT TRANSMEM 3044 3064
 FT TRANSMEM 3065 3085
 FT TRANSMEM 3086 3106
 FT TRANSMEM 3107 3127
 FT TRANSMEM 3128 3148
 FT TRANSMEM 3149 3169
 FT TRANSMEM 3170 3190
 FT TRANSMEM 3191 3210
 FT TRANSMEM 3211 3231
 FT TRANSMEM 3232 3252
 FT TRANSMEM 3253 3272
 FT TRANSMEM 3273 3293
 FT TRANSMEM 3294 3314
 FT TRANSMEM 3315 3335
 FT TRANSMEM 3336 3356
 FT TRANSMEM 3357 3377
 FT TRANSMEM 3378 3398
 FT TRANSMEM 3399 3419
 FT TRANSMEM 3420 3440
 FT TRANSMEM 3441 3461
 FT TRANSMEM 3462 3482
 FT TRANSMEM 3483 3503
 FT TRANSMEM 3504 3524
 FT TRANSMEM 3525 3545
 FT TRANSMEM 3546 3566
 FT TRANSMEM 3567 3587
 FT TRANSMEM 3588 3608
 FT TRANSMEM 3609 3629
 FT TRANSMEM 3630 3650
 FT TRANSMEM 3651 3671
 FT TRANSMEM 3672 3692
 FT TRANSMEM 3693 3713
 FT TRANSMEM 3714 3734
 FT TRANSMEM 3735 3755
 FT TRANSMEM 3756 3776
 FT TRANSMEM 3777 3797
 FT TRANSMEM 3798 3818
 FT TRANSMEM 3819 3839
 FT TRANSMEM 3840 3860
 FT TRANSMEM 3861 3881
 FT TRANSMEM 3882 3902
 FT TRANSMEM 3903 3923
 FT TRANSMEM 3924 3944
 FT TRANSMEM 3945 3965
 FT TRANSMEM 3966 3986
 FT TRANSMEM 3987 4007
 FT TRANSMEM 4008 4028
 FT TRANSMEM 4029 4049
 FT TRANSMEM 4050 4070
 FT TRANSMEM 4071 4091
 FT TRANSMEM 4092 4112
 FT TRANSMEM 4113 4133
 FT TRANSMEM 4134 4154
 FT TRANSMEM 4155 4175
 FT TRANSMEM 4176 4196
 FT TRANSMEM 4197 4217
 FT TRANSMEM 4218 4238
 FT TRANSMEM 4239 4259
 FT TRANSMEM 4260 4280
 FT TRANSMEM 4281 4301
 FT TRANSMEM 4302 4322
 FT TRANSMEM 4323 4343
 FT TRANSMEM 4344 4364
 FT TRANSMEM 4365 4385
 FT TRANSMEM 4386 4406
 FT TRANSMEM 4407 4427
 FT TRANSMEM 4428 4448
 FT TRANSMEM 4449 4469
 FT TRANSMEM 4470 4490
 FT TRANSMEM 4491 4511
 FT TRANSMEM 4512 4532
 FT TRANSMEM 4533 4553
 FT TRANSMEM 4554 4574
 FT TRANSMEM 4575 4595
 FT TRANSMEM 4596 4616
 FT TRANSMEM 4617 4637
 FT TRANSMEM 4638 4658
 FT TRANSMEM 4659 4679
 FT TRANSMEM 4680 4700
 FT TRANSMEM 4701 4721
 FT TRANSMEM 4722 4742
 FT TRANSMEM 4743 4763
 FT TRANSMEM 4764 4784
 FT TRANSMEM 4785 4805
 FT TRANSMEM 4806 4826
 FT TRANSMEM 4827 4847
 FT TRANSMEM 4848 4868
 FT TRANSMEM 4869 4889
 FT TRANSMEM 4890 4910
 FT TRANSMEM 4911 4931
 FT TRANSMEM 4932 4952
 FT TRANSMEM 4953 4973
 FT TRANSMEM 4974 4994
 FT TRANSMEM 4995 5015
 FT TRANSMEM 5016 5036
 FT TRANSMEM 5037 5057
 FT TRANSMEM 5058 5078
 FT TRANSMEM 5079 5099
 FT TRANSMEM 5100 5120
 FT TRANSMEM 5121 5141
 FT TRANSMEM 5142 5162
 FT TRANSMEM 5163 5183
 FT TRANSMEM 5184 5204
 FT TRANSMEM 5205 5225
 FT TRANSMEM 5226 5246
 FT TRANSMEM 5247 5267
 FT TRANSMEM 5268 5288
 FT TRANSMEM 5289 5309
 FT TRANSMEM 5310 5330
 FT TRANSMEM 5331 5351
 FT TRANSMEM 5352 5372
 FT TRANSMEM 5373 5393
 FT TRANSMEM 5394 5414
 FT TRANSMEM 5415 5435
 FT TRANSMEM 5436 5456
 FT TRANSMEM 5457 5477
 FT TRANSMEM 5478 5498
 FT TRANSMEM 5499 5519
 FT TRANSMEM 5520 5540
 FT TRANSMEM 5541 5561
 FT TRANSMEM 5562 5582
 FT TRANSMEM 5583 5603
 FT TRANSMEM 5604 5624
 FT TRANSMEM 5625 5645
 FT TRANSMEM 5646 5666
 FT TRANSMEM 5667 5687
 FT TRANSMEM 5688 5708
 FT TRANSMEM 5709 5729
 FT TRANSMEM 5730 5750
 FT TRANSMEM 5751 5771
 FT TRANSMEM 5772 5792
 FT TRANSMEM 5793 5813
 FT TRANSMEM 5814 5834
 FT TRANSMEM 5835 5855
 FT TRANSMEM 5856 5876
 FT TRANSMEM 5877 5897
 FT TRANSMEM 5898 5918
 FT TRANSMEM 5919 5939
 FT TRANSMEM 5940 5960
 FT TRANSMEM 5961 5981
 FT TRANSMEM 5982 6002
 FT TRANSMEM 6003 6023
 FT TRANSMEM 6024 6044
 FT TRANSMEM 6045 6065
 FT TRANSMEM 6066 6086
 FT TRANSMEM 6087 6107
 FT TRANSMEM 6108 6128
 FT TRANSMEM 6129 6149
 FT TRANSMEM 6150 6170
 FT TRANSMEM 6171 6191
 FT TRANSMEM 6192 6212
 FT TRANSMEM 6213 6232
 FT TRANSMEM 6233 6253
 FT TRANSMEM 6254 6274
 FT TRANSMEM 6275 6295
 FT TRANSMEM 6296 6316
 FT TRANSMEM 6317 6337
 FT TRANSMEM 6338 6358
 FT TRANSMEM 6359 6379
 FT TRANSMEM 6380 6400
 FT TRANSMEM 6401 6421
 FT TRANSMEM 6422 6442
 FT TRANSMEM 6443 6463
 FT TRANSMEM 6464 6484
 FT TRANSMEM 6485 6505
 FT TRANSMEM 6506 6526
 FT TRANSMEM 6527 6547
 FT TRANSMEM 6548 6568
 FT TRANSMEM 6569 6589
 FT TRANSMEM 6590 6610
 FT TRANSMEM 6611 6630
 FT TRANSMEM 6631 6651
 FT TRANSMEM 6652 6672
 FT TRANSMEM 6673 6693
 FT TRANSMEM 6694 6714
 FT TRANSMEM 6715 6735
 FT TRANSMEM 6736 6756
 FT TRANSMEM 6757 6777
 FT TRANSMEM 6778 6798
 FT TRANSMEM 6799 6819
 FT TRANSMEM 6820 6840
 FT TRANSMEM 6841 6861
 FT TRANSMEM 6862 6882
 FT TRANSMEM 6883 6903
 FT TRANSMEM 6904 6924
 FT TRANSMEM 6925 6945
 FT TRANSMEM 6946 6966
 FT TRANSMEM 6967 6987
 FT TRANSMEM 6988 7008
 FT TRANSMEM 7009 7029
 FT TRANSMEM 7030 7050
 FT TRANSMEM 7051 7071
 FT TRANSMEM 7072 7092
 FT TRANSMEM 7093 7113
 FT TRANSMEM 7114 7134
 FT TRANSMEM 7135 7155
 FT TRANSMEM 7156 7176
 FT TRANSMEM 7177 7197
 FT TRANSMEM 7198 7218
 FT TRANSMEM 7219 7239
 FT TRANSMEM 7240 7260
 FT TRANSMEM 7261 7281
 FT TRANSMEM 7282 7302
 FT TRANSMEM 7303 7323
 FT TRANSMEM 7324 7344
 FT TRANSMEM 7345 7365
 FT TRANSMEM 7366 7386
 FT TRANSMEM 7387 7407
 FT TRANSMEM 7408 7428
 FT TRANSMEM 7429 7449
 FT TRANSMEM 7450 7470
 FT TRANSMEM 7471 7491
 FT TRANSMEM 7492 7512
 FT TRANSMEM 7513 7533
 FT TRANSMEM 7534 7554
 FT TRANSMEM 7555 7575
 FT TRANSMEM 7576 7596
 FT TRANSMEM 7597 7617
 FT TRANSMEM 7618 7638
 FT TRANSMEM 7639 7659
 FT TRANSMEM 7660 7680
 FT TRANSMEM 7681 7701
 FT TRANSMEM 7702 7722
 FT TRANSMEM 7723 7743
 FT TRANSMEM 7744 7764
 FT TRANSMEM 7765 7785
 FT TRANSMEM 7786 7806
 FT TRANSMEM 7807 7827
 FT TRANSMEM 7828 7848
 FT TRANSMEM 7849 7869
 FT TRANSMEM 7870 7890
 FT TRANSMEM 7891 7911
 FT TRANSMEM 7912 7932
 FT TRANSMEM 7933 7953
 FT TRANSMEM 7954 7974
 FT TRANSMEM 7975 7995
 FT TRANSMEM 7996 8016
 FT TRANSMEM 8017 8037
 FT TRANSMEM 8038 8058
 FT TRANSMEM 8059 8079
 FT TRANSMEM 8080 8100
 FT TRANSMEM 8101 8121
 FT TRANSMEM 8122 8142
 FT TRANSMEM 8143 8163
 FT TRANSMEM 8164 8184
 FT TRANSMEM 8185 8205
 FT TRANSMEM 8206 8226
 FT TRANSMEM 8227 8247
 FT TRANSMEM 8248 8268
 FT TRANSMEM 8269 8289
 FT TRANSMEM 8290 8310
 FT TRANSMEM 8311 8331
 FT TRANSMEM 8332 8352
 FT TRANSMEM 8353 8373
 FT TRANSMEM 8374 8394
 FT TRANSMEM 8395 8415
 FT TRANSMEM 8416 8436
 FT TRANSMEM 8437 8457
 FT TRANSMEM 8458 8478
 FT TRANSMEM 8479 8499
 FT TRANSMEM 8500 8520
 FT TRANSMEM 8521 8541
 FT TRANSMEM 8542 8562
 FT TRANSMEM 8563 8583
 FT TRANSMEM 8584 8604
 FT TRANSMEM 8605 8625
 FT TRANSMEM 8626 8646
 FT TRANSMEM 8647 8667
 FT TRANSMEM 8668 8688
 FT TRANSMEM 8689 8709
 FT TRANSMEM 8710 8730
 FT TRANSMEM 8731 8751
 FT TRANSMEM 8752 8772
 FT TRANSMEM 8773 8793
 FT TRANSMEM 8794 8814
 FT TRANSMEM 8815 8835
 FT TRANSMEM 8836 8856
 FT TRANSMEM 8857 8877
 FT TRANSMEM 8878 8898
 FT TRANSMEM 8899 8919
 FT TRANSMEM 8920 8940
 FT TRANSMEM 8941 8961
 FT TRANSMEM 8962 8982
 FT TRANSMEM 8983 9003
 FT TRANSMEM 9004 9024
 FT TRANSMEM 9025 9045
 FT TRANSMEM 9046 9066
 FT TRANSMEM 9067 9087
 FT TRANSMEM 9088 9108
 FT TRANSMEM 9109 9129
 FT TRANSMEM 9130 9150
 FT TRANSMEM 9151 9171
 FT TRANSMEM 9172 9192
 FT TRANSMEM 9193 9213
 FT TRANSMEM 9214 9234
 FT TRANSMEM 9235 9255
 FT TRANSMEM 9256 9276
 FT TRANSMEM 9277 9297
 FT TRANSMEM 9298 9318
 FT TRANSMEM 9319 9339
 FT TRANSMEM 9340 9360
 FT TRANSMEM 9361 9381
 FT TRANSMEM 9382 9402
 FT TRANSMEM 9403 9423
 FT TRANSMEM 9424 9444
 FT TRANSMEM 9445 9465
 FT TRANSMEM 9466 9486
 FT TRANSMEM 9487 9507
 FT TRANSMEM 9508 9528
 FT TRANSMEM 9529 9549
 FT TRANSMEM 9550 9570
 FT TRANSMEM 9571 9591
 FT TRANSMEM 9592 9612
 FT TRANSMEM 9613 9633
 FT TRANSMEM 9634 9654
 FT TRANSMEM 9655 9675
 FT TRANSMEM 9676 9696
 FT TRANSMEM 9697 9717
 FT TRANSMEM 9718 9738
 FT TRANSMEM 9739 9759
 FT TRANSMEM 9760 9780
 FT TRANSMEM 9781 9801
 FT TRANSMEM 9802 9822
 FT TRANSMEM 9823 9843
 FT TRANSMEM 9844 9864
 FT TRANSMEM 9865 9885
 FT TRANSMEM 9886 9906
 FT TRANSMEM 9907 9927
 FT TRANSMEM 9928 9948
 FT TRANSMEM 9949 9969
 FT TRANSMEM 9970 9990
 FT TRANSMEM 9991 10000
 FT TRANSMEM 10001 10021
 FT TRANSMEM 10022 10043
 FT TRANSMEM 10044 10064
 FT TRANSMEM 10065 10085
 FT TRANSMEM 10086 10106
 FT TRANSMEM 10107 10127
 FT TRANSMEM 10128 10149
 FT TRANSMEM 10150 10170
 FT TRANSMEM 10171 10191
 FT TRANSMEM 10192 10212
 FT TRANSMEM 10213 10233
 FT TRANSMEM 10234 10254
 FT TRANSMEM 10255 10275
 FT TRANSMEM 10276 10296
 FT TRANSMEM 10297 10317
 FT TRANSMEM 10318 10338
 FT TRANSMEM 10339 10359
 FT TRANSMEM 10360 10380
 FT TRANSMEM 10381 10401
 FT TRANSMEM 10402 10422
 FT TRANSMEM 10423 10443
 FT TRANSMEM 10444 10464
 FT TRANSMEM 10465 10485
 FT TRANSMEM 10486 10506
 FT TRANSMEM 10507 10527
 FT TRANSMEM 10528 10548
 FT TRANSMEM 10549 10569
 FT TRANSMEM 10570 10590
 FT TRANSMEM 10591 10611
 FT TRANSMEM 10612 10632
 FT TRANSMEM 10633 10653
 FT TRANSMEM 10654 10674
 FT TRANSMEM 10675 10695
 FT TRANSMEM 10696 10716
 FT TRANSMEM 10717 10737
 FT TRANSMEM 10738 10758
 FT TRANSMEM 10759 10779
 FT TRANSMEM 10780 10800
 FT TRANSMEM 10801 10821
 FT TRANSMEM 10822 10842
 FT TRANSMEM 10843 10863
 FT TRANSMEM 10864 10884
 FT TRANSMEM 10885 10905
 FT TRANSMEM 10906 10926
 FT TRANSMEM 10927 10947
 FT TRANSMEM 10948 10968
 FT TRANSMEM 10969 10989
 FT TRANSMEM 10990 11010
 FT TRANSMEM 11011 11031
 FT TRANSMEM 11032 11052
 FT TRANSMEM 11053 11073
 FT TRANSMEM 11074 11094
 FT TRANSMEM 11095 11115
 FT TRANSMEM 11116 11136
 FT TRANSMEM 11137 11157
 FT TRANSMEM 11158 11178
 FT TRANSMEM 11179 11200
 FT TRANSMEM 11201 11221
 FT TRANSMEM 11222 11242
 FT TRANSMEM 11243 11263
 FT TRANSMEM 11264 11284
 FT TRANSMEM 11285 11305
 FT TRANSMEM 11306 11326
 FT TRANSMEM 11327 11347
 FT TRANSMEM 11348 11368
 FT TRANSMEM 11369 11389
 FT TRANSMEM 11390 11410
 FT TRANSMEM 11411 11431
 FT TRANSMEM 11432 11452
 FT TRANSMEM 11453 11473
 FT TRANSMEM 11474 11494
 FT TRANSMEM 11495 11515
 FT TRANSMEM 11516 11536
 FT TRANSMEM 11537 11557
 FT TRANSMEM 11558 11578
 FT TRANSMEM 11579 11599
 FT TRANSMEM 11600 11620
 FT TRANSMEM 11621 11641
 FT TRANSMEM 11642 11662
 FT TRANSMEM 11663 11683
 FT TRANSMEM 11684 11704
 FT TRANSMEM 11705 11725
 FT TRANSMEM 11726 11746
 FT TRANSMEM 11747 11767
 FT TRANSMEM 11768 11788
 FT TRANSMEM 11789 11809
 FT TRANSMEM 11810 11830
 FT TRANSMEM 11831 11851
 FT TRANSMEM 11852 11872
 FT TRANSMEM 11873 11893
 FT TRANSMEM 11894 11914
 FT TRANSMEM 11915 11935
 FT TRANSMEM 11936 11956
 FT TRANSMEM 11957 11977
 FT TRANSMEM 11978 11998
 FT TRANSMEM 11999 12019
 FT TRANSMEM 12020 12040
 FT TRANSMEM 12041 12061
 FT TRANSMEM 12062 12082
 FT TRANSMEM 12083 12103
 FT TRANSMEM 12104 12124
 FT TRANSMEM 12125 12145
 FT TRANSMEM 12146 12166
 FT TRANSMEM 12167 12187
 FT TRANSMEM 12188 12208
 FT TRANSMEM 12209 12229
 FT TRANSMEM 12230 12250
 FT TRANSMEM 12251 12271
 FT TRANSMEM 12272 12292
 FT TRANSMEM 12293 12313
 FT TRANSMEM 12314 12334
 FT TRANSMEM 12335 12355
 FT TRANSMEM 12356 12376
 FT TRANSMEM 12377 12397
 FT TRANSMEM 12398 12418
 FT TRANSMEM 12419 12439
 FT TRANSMEM 12440 12460
 FT TRANSMEM 12461 12481
 FT TRANSMEM 12482 12502
 FT TRANSMEM 12503 12523
 FT TRANSMEM 12524 12544
 FT TRANSMEM 12545 12565
 FT TRANSMEM 12566 12586
 FT TRANSMEM 12587 12607
 FT TRANSMEM 12608 12628
 FT TRANSMEM 12629 12649
 FT TRANSMEM 12650 12670
 FT TRANSMEM 12671 12691
 FT TRANSMEM 12692 12712
 FT TRANSMEM 12713 12733
 FT TRANSMEM 12734 12754
 FT TRANSMEM 12755 12775
 FT TRANSMEM 12776 12796
 FT TRANSMEM 12797 12817
 FT TRANSMEM 12818 12838
 FT TRANSMEM 12839 12859
 FT TRANSMEM 12860 12880
 FT TRANSMEM 12881 12901
 FT TRANSMEM 12902 12922
 FT TRANSMEM 12923 12943
 FT TRANSMEM 12944 12964
 FT TRANSMEM 12965 12985
 FT TRANSMEM 12986 13006
 FT TRANSMEM 13007 13027
 FT TRANSMEM 13028 13048
 FT TRANSMEM 13049 13069
 FT TRANSMEM 13070 13090
 FT TRANSMEM 13091 13111
 FT TRANSMEM 13112 13132
 FT TRANSMEM 13133 13153
 FT TRANSMEM 13154 13174
 FT TRANSMEM 13175 13195
 FT TRANSMEM 13196 13216
 FT TRANSMEM 13217 13237
 FT TRANSMEM 13238 13258
 FT TRANSMEM 13259 13279
 FT TRANSMEM 13280 13300
 FT TRANSMEM 13301 13321
 FT TRANSMEM 13322 13342
 FT TRANSMEM 13343 13363
 FT TRANSMEM 13364 13384
 FT TRANSMEM 13385 13405
 FT TRANSMEM 13406 13426
 FT TRANSMEM 13427 13447
 FT TRANSMEM 13448 13468
 FT TRANSMEM 13469 13489
 FT TRANSMEM 13490 13510
 FT TRANSMEM 13511 13531
 FT TRANSMEM 13532 13552
 FT TRANSMEM 13553 13573
 FT TRANSMEM 13574 13594
 FT TRANSMEM 13595 13615
 FT TRANSMEM 13616 13636
 FT TRANSMEM 13637 13657
 FT TRANSMEM 13658 13678
 FT TRANSMEM 13679 13699
 FT TRANSMEM 13700 13720
 FT TRANSMEM 13721 13741
 FT TRANSMEM 13742 13762
 FT TRANSMEM 13763 13783
 FT TRANSMEM 13784 13804
 FT TRANSMEM 13805 13825
 FT TRANSMEM 13826 13846
 FT TRANSMEM 13847 13867
 FT TRANSMEM 13868 13888
 FT TRANSMEM 13889 13909
 FT TRANSMEM 13910 13930
 FT TRANSMEM 13931 13951
 FT TRANSMEM 13952 13972
 FT TRANSMEM 13973 13993
 FT TRANSMEM 13994 14014
 FT TRANSMEM 14015 14035
 FT TRANSMEM 14036 14056
 FT TRANSMEM 14057 14077
 FT TRANSMEM 14078 14098
 FT TRANSMEM 14099 14119
 FT TRANSMEM 14120 14140
 FT TRANSMEM 14141 14161
 FT TRANSMEM 14162 14182
 FT TRANSMEM 14183 14203
 FT TRANSMEM 14204 1

CC	modified and this statement is not removed. Usage by and for commercial
CC	entitles requires a license agreement (see http://www.isb-slb.ch/announce/).
CC	or send an email to license@isb-slb.ch .
CC	-----
DR	EMBL; M34503; AAA2474.1; -
DR	PIR; J00420; J00420.
DR	InterPro: IPR000757; Glyco_hydro_16.
DR	Pfam; PF00722; Glyco_hydro_16.1.
DR	ProSITE; PS01034; GLYCOSYL_HYDROL_F16.1.
DR	Cell wall; Hydrolase; Glycosidase; Signal.
FT	SIGNAL
FT	CHAIN 1 38
FT	ACT_SITE 39 682 GLUCAN ENDO-1,3-BETA-GLUCOSIDASE A1.
FT	ACT_SITE 552 552 NUCLEOPHILE (BY SIMILARITY).
FT	ACT_SITE 557 557 PROTON DONOR (BY SIMILARITY).
SO	SEQUENCE 682 AA: 75465 MW: 8045470E34DAD4B5 CRC64;
Query Match	3.2%; Score 142; DB 1; Length 682;
Best Local Similarity	19.8%; Pred. No. 0.29;
Matches	114; Conservative 82; Mismatches 197; Indels 184; Gaps 32;
DB	308 SOEHHDMIRGELASGRPYVY-AGNNQSIGAFVCCGYASDGTFRHNMW-WG-GVSNQGYK 364
	78 SATMND-----YSDVGVAVKGNMNMWIDIDAGGYIYNQMGHMSDGGPNGYW- 125
OY	365 LTLISPSLIGGSGIGF-----TLYQDI-----TGIRPATPAAGDAPLALAKRIE 415
DB	126 FTLSATTEITOLYSKANGKLEYQLVFNQINKTTITPAAMPQGP-----QIT 171
OY	416 AEKYESGSLANGSYIYNTGEEQSNLDLGYRLNKADGEIVETKSSIN-ISMWYGEHPES 474
DB	172 ASFLGAGFT--YLPFNNDASVT-----YEAAADLAKYVKKRYVNSSMIDIDNNAA 221
OY	475 FSLAPNOLSGCINMTTLLYRRTGTEQWEPVRAAGGYNSIKVNTDNNVYVVDNNEG 534
DB	222 GMIYDHNFGQ-----FTDGGGGYFNVTES---INVKLESKTSANLVIYTFENEP 269
OY	535 KLSIVPNSFVADLNSYEHSHTTTCVQENSPPDEIRTPVAFALSTGATADVISLGVMAV 594
DB	270 -----TRNSIV--IIPYEGTITPADANG-----SIGLPKRI 299
OY	595 PGSSSNPVVWSKDVLLTSEGDYTLWYRFSINNQDEMKIGISVSKPTETHTPLFEVG 654
DB	300 DGA--PI--AKEL-----GNFY--YQININGQ--W-----VLSNSSQSKFYASNG 339
OY	655 HNQIST-----YTLDMANH---RYLPDTLKNLGLPFNGE-----LVVFRQTOSS 697
DB	340 YNNSDANQOMGYMADYIYGLWFQPIQENMQIRIGYPLNGQAGNIGNFNVTYFIGNPA 399
OY	698 SGSLMAQGEYHIK-----QGEFFYKPVVEGPIP-----GSY----- 731
DB	400 PRPDVSDDEDISIGTPDPAIAGNMLIWQDEFNGTTLDTSKWNYETGYLLANDPATWGW 459
OY	732 RATLHAEPNGQOOYLK-GKRNYVYKIVNGTAVEAIESSSEIRVPPNPAADVETASPCI 790
DB	460 NAELQHYNSQNYVYQDKLN-----IKAMDSKSPQDPNNRYAQIS----- 502
OY	791 PQTSTIIFDLGSKIYMKNSLSAGHRADV--SRLPNG 826
DB	503 -----SGKINTKDKLSIKYGRVDFRAKPLPTG 528
RESULT 9	
KEX2_YEAST	STANDARD; PRT; 814 AA.
AC	PI3134;
DT	01-JAN-1990 (Rel. 13, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DE	Kexin precursor (EC 3.4.21.61) (KEX2 protease) (Proteinase YSCF).
GN	KEX2 OR OPS1 OR YNL238W OR N1122.
OS	Saccharomyces cerevisiae (Baker's yeast).
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=89025851; PubMed=2845974;
RA Mizuno K., Nakamura T., Ohshima T., Tanaka S., Matsuo H.;
RT "Yeast Kex2 genes encodes an endopeptidase homologous to
RL subtilisin-like serine proteases.";
RN Biochem. Biophys. Res. Commun. 156:246-254(1988).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=89160792; PubMed=2646633;
RA Fuller R.S., Brake A., Thorner J.;
RT "Yeast prothormone processing enzyme (KEX2 gene product) is a Ca2+-
RL proc. Natl. Acad. Sci. U.S.A. 86:1434-1438(1989).
RN [3]
RN SEQUENCE FROM N.A.
RX MEDLINE=97051596; PubMed=8896273;
RA Pandolfo D., de Antoni A., Lanfranchi G., Valle G.;
RT "The DNA sequence of cosmid 14-5 from chromosome XIV reveals 21 open
RL reading frames including a novel gene encoding a globin-like
RN domain.";
RN Yeast 12:1071-1076(1996).
RN [4]
RN PROCESSING, AND CHARACTERIZATION.
RX MEDLINE=92141230; PubMed=1736307;
RA Brenner C., Fuller R.S.;
RT "Structural and enzymatic characterization of a purified prothormone-
RL processing enzyme: secreted, soluble Kex2 protease.";
RN Proc. Natl. Acad. Sci. U.S.A. 89:922-926(1992).
RN [5]
RN PROCESSING.
RX MEDLINE=92183954; PubMed=1544507;
RA Germain D., Dumas F., Vernet T., Bourbonnais Y., Thomas D.Y.,
RL Bollen G.;
RT "The pro-region of the Kex2 endoprotease of *Saccharomyces cerevisiae*
is removed by self-processing.";
RN FEBS Lett. 299:283-286(1992).
RN [6]
RN POST-TRANSLATIONAL MODIFICATIONS.
RX MEDLINE=92011876; PubMed=1918142;
RA Wilcox C.A., Fuller R.S.;
RT "Posttranslational processing of the prothormone-cleaving Kex2
RL protease in the *Saccharomyces cerevisiae* secretory pathway.";
RN J. Cell Biol. 115:297-307(1991).
RN [7]
RN FUNCTION: PROCESSING OF PRECURSORS OF ALPHA-FACTORS AND KILLER
RL TOXIN.
RN [8]
RN CATALYTIC ACTIVITY: Cleavage of Lys-Arg-|-Xaa and Arg-Arg-|-Xaa
RN bonds to process yeast alpha-factor prothormone and killer toxin
RN precursors.
RN [9]
RN COFACTOR: CALCIUM DEPENDENT.
RN [10]
RN SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. LATE GOLGI
RN COMPARTMENT.
RN [11]
RN PTM: O-GlcCOSYLATED.
RN [12]
RN SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8, FURIN SUBFAMILY.
RN [13]
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
RN [14]
RN EMBL, M22870; AAA34719.1; -
RN EMBL, M24201; AAA34718.1; -
RN EMBL, Z69381; CA93360.1; -
RN EMBL, Z71514; CA96143.1; -
RN PIR, A28931; KXBY.
RN PIR, S42157; S42157.
RN HSSP, P04072; 1THM.
RN MEROPS, S08.070; -
RN SGD; S0005182; KEX2.

Page 9

DR InterPro: IPR002884; P_domain.
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR Pfam: PF01483; P_PARTIAL.
 DR PRINTS: PRO00723; SUBTILISIN.
 DR PRODOM: PD000717; P_domain; 1.
 DR PROSITE: PS00136; SUBTILASE_ASP; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 KM Hydroxylase: Serine protease; Glycoprotein; Calcium; Transmembrane;
 KM Zymogen; Signal.
 FT SIGNAL 1 19
 FT PROPEP 20 109
 FT PROPEP 110 113
 FT CHAIN 114 814
 FT DOMAIN 114 678
 FT TRANSMEM 679 699
 FT DOMAIN 700 814
 FT DOMAIN 152 410
 FT DOMAIN 625 642
 FT ACT_SITE 175 175
 FT ACT_SITE 213 213
 FT ACT_SITE 385 385
 FT CARBOHYD 42 42
 FT CARBOHYD 163 163
 FT CARBOHYD 404 404
 FT CARBOHYD 480 480
 SO SEQUENCE 814 AA; 90002 MW; 5DCEEDADCF439C78 CRC64;

Query Match

Best Local Similarity 3.18; Score 139.5; DB 1; Length 814;
 Matches 167; Conservative 80; Mismatches 278; Indels 231; Gaps 43;

QY 166 NDPRNNGYPMNKPRLPN-----GNHATGCVATAQAQIMRHSMPLC-----211
 DB 134 NDPLFRQ---WLVNPSFPGSDINVLDMYNNITGAGVAAAYDDGLDENDLNDNC 190
 QY 212 GEGSFYHAGSLVGNNGSTFGEMYDINMGPNDDLNLQSOVDAYAT-LMRDVSASVM 270
 DB 191 AEGSMDFN-----DMTNLP-KPRLSD-----DYHGTROAGEIAAKGW 227
 QY 271 SFENSGSYVYVGCALRNFRKRSLO-----LHVALYTSQEM-----311
 DB 228 NFGCVGVG-YNAKISG-----RLSGDITTEDEASLGLDVNDIY-SCSWGPAD 277
 QY 312 -----IDMIRGEIAGSPRYXAGNNGSIGHAFCVCDGAS--DGTFFHMGCGVSN 361
 DB 278 DGRHLOGPSDLYKALVKG-----VTGGRSKGAIIV---FASNGGTGDCNCTYDGTNS 330
 QY 362 FYKLT-----LISPTSLGIG-----GEGIGFTYQEIITGI-----EPAKTPA 399
 DB 331 IYSITIGAIIDHKDLHPYSEGCASVMAVYSSGSEYIHSSDINGCSNHSHTSAAPL 390
 QY 400 EAGT-----DALPILAKDIEAFYKSESLNGCYSTYNGEBSNIDLGRLKAKAGEVI 454
 DB 391 AAGVYTLLEANNLNRDVOY-----LSTLSAVGLEKNADGDMR-DSANGKRY 438
 QY 455 EVYTSINISWYGEHPESFSLAPNQSOGINTITLLRRGTGEQMEPRHAQGVNS 514
 DB 439 SHR-----YFGGIDAHKLIEMSKTWNENVAOWEFLPT-----LYVSO 477
 QY 515 IKVNTDPNNVVYTNNEGKLSIVNFSVADLNSYFTTVOFNSDSPDEIRTPVAPA 574
 DB 478 STNSTRETTLESVITI--SEKSLQ-----DANFKRIEHTVIVY-----DIDTEIR 519
 QY 575 LSTGATADVYISLGMVAEYVPGG--SSNYVYWSKVDLITSEGDYITMYRFSI-----NNOK 629
 DB 520 -----GTTTVDLIS-----PAQIISNLGVVRPRD--SSEG-FKXMTWMSVANGENY 565
 QY 630 DEMKRTIGSVKTPREYTH-----PLF--EVGHNOTSTYTLDMANHRVLPDPTLKN 678
 DB 566 GGMK-----IKVKT--TENGHRIDFHSWRLKLFGEISDSKTEITFVFNOKDEVEPAATTEST 620

QY 679 LGIPFENGELVWVFRQTOSSSGSLMAOETVHIKOGETFYKPYVEGPID--GSYRATLH 736
 DB 621 VQSYASSTISISATSTSSISIT--GVEISALPDQTTASTDPDSDNPTFKLSSPQAMH 678
 QY 737 AFVN-----GOOOL-----YLKGRNYTYKYNGTAVAEI-----SSEETRVPPNPAR 780
 DB 679 YFTLPIGATFATLVLYFMFEMKSRRR-----IRSRFAEYERDIDTDEYSDTLNDGYS 733
 QY 781 DVEISAPCIPQETSIILDLGKIYMKNSLSAGH 816
 DB 734 GITE-----PEVEEDPFDLSDEHDHLASSSENG 763

RESULT 10

BCNS_CLOPE
 ID BCNS_CLOPE STANDARD; PRT; 890 AA.

AC P08656;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE Bacteriocin BCNS.
 GN BCN.
 OS Clostridium perfringens.
 OG Plasmid pIP404.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_Taxid=1502;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CPN50;
 RX MEDLINE=87057020; PubMed=2901768;
 RA Garner T., Cole S.T.;
 RT "Complete nucleotide sequence and genetic organization of the
 RT bacteriocinogenic plasmid, pIP404, from Clostridium perfringens.";
 RL Plasmid 19:134-150(1988).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CPN50;
 RX MEDLINE=87057020; PubMed=2877971;
 RA Garner T., Cole S.T.;
 RT "Characterization of a bacteriocinogenic plasmid from Clostridium
 RT perfringens and molecular genetic analysis of the
 RT bacteriocin-encoding gene.";
 RL J. Bacteriol. 168:1189-1196(1986).
 RN (3)
 RP SEQUENCE OF 1-14 FROM N.A.
 RC STRAIN=CPN50;
 RX MEDLINE=89039249; PubMed=2460717;
 RA Garner T., Cole S.T.;
 RT "Studies of UV-inducible promoters from Clostridium perfringens in
 RT vivo and in vitro.";
 RL MOL. Microbiol. 2:607-614(1988).
 CC -1- FUNCTION: MAY FUNCTION AS AN IONOPHORE.
 CC -1- INDUCTION: BY UV IRRADIATION.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: M14481; AAA98248.1;
 DR EMBL: M32882; AAA98249.1;
 DR PIR: A30481; A30481.
 DR InterPro: IPR003646; SH3_bac.
 DR SMART: SM00287; SH3b; 3.
 KM Antibiotic; Bacteriocin; Plasmid.
 FT DOMAIN 815 869
 FT SEQUENCE 890 AA; 96699 MW; F4E5E8971C31C6C6 CRC64;

Query Match

3.18; Score 139.5; DB 1; Length 890;

Best Local Similarity 19.5%; Pred. No. 0.6;
Matches 152; Conservative 95; Mismatches 240; Indels 291; Gaps 41;

QY 142 NEELRLRBE-----VAEVAHLM--DNG--HFPANDPRMNOGTPMKNKEPLRNGHAY 191
Db 258 SEEVLDENGGHSLNPEYEAATLLERKMKHVVYD-----TNKGPNTSKSGYKY 307
QY 192 TGVAFATA-----AOIMRY-----HSMPLQ----- 211
Db 308 EGAAATRYDIPPSITNAQIVYIGSGRGLAAIKYKNGNSLVEFCAIHGMEDNNMAD 367
QY 212-----GEGSFYDHAAGSLVGNMGTGEMTWMNMGNDPLDMLTQSDVAYATLMRDV 264
Db 368 GIELTRIGNGLIEHFQNGNTNNS-----LY--IIPVANDP----- 401
QY 265 SASVSASFENGSGTYSYVVGALRNNF-----RKYSLLQHLRALYTSQEMH 312
Db 402--GISEPTNNGPRCT--IYGAVDNDNDFLPGSPGVPRYHSGSE--PLVSSEKSLH 455
QY 313 DMIGELASGRPVYAGNNOGIGHAFVCDGYASDGTFFHNMGMGVSNNGFYKLTLLSPTS 372
Db 456 DFICGV-----KNRTSGEMCYD-----LH--GMEGAALGNPEIGEFYFNO 494
QY 373 LGIG-----GEGIFTIYOEIITIGIEPAKTPAEGTDLPLTALKDIEAEKSSBSGLNVG 427
Db 495 FGFGORSGYGDNRGFMIGMAKSGAK-----AALIELPGSTKSHSDVYNG 539
QY 428 YSIYNTGEGSNL-----DLGYRLNKADGEYIEVKTSSINISWYGYGHP 472
Db 540 RYLQKIINAVNVLIGSGSSSSGSSFSFDSVY--EATGEVINO--SFLNVR--EAGLYT 594
QY 473 ESFLAPNOLSGQINTITILYRTGTGEQEVPRH--AOGYNSIKVNTDNNVYVDN 531
Db 595 NSI-----GQLNG--NKVNIYAK--NGDMYKIKGSEGYNS----- 629
QY 532 NEGLSTVPSNFVADLNSYEHSTTVQFNSDPDEIRTPVAFALSTGATADVLSLWVM 591
Db 630--GYIILKNTSVKLEDMQ-----EDCIRFGM-- 655
QY 592 AVEGCGSSNPVWSKDLVLTSEBGDTYLMRFSTINOKDEKKIGSVSKPTPE----- 646
Db 656-----GPITKEK--YLEYMDSTLYKSIENDISQAIKNSLINVPLNFSVSEM 703
QY 647--THPLFEVGNHNOTSTYTLDMANRPLDFTLKNLGPENGELVYVFR-----QT 694
Db 704 IACQIYV--NNETTSFRDEWYSKSNPNFIVKTKLS--NGQIIVDRINIKREKIKT 759
QY 695 QSSSGSLMAOETVHIKOGTFYKPV-----VEGPIDGSTRATLHAFVNGOQOYLK 748
Db 760 KIPRAKGAFKDTIFE-----FFKIGDGMFTALSGAISGDSLVFQS--NGE--LK 808
QY 749 GKRNTYVIVGTAVEALESSEIRVFPNPADYIEISAPCIPOTSITLIDLSKTY 806
Db 809 SNEDIKALAAVINGVET-----MFCALIGFI--AOCIAPEPIVA--AVAGATV 857

RESULT 11
PLS_PYREFU STANDARD: PRT: 1398 AA.
AC P72186;
DT 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DN Pyrolysin precursor (EC 3.4.21.-).
CN PLS OR PF0287.
OS Pyrococcus furiosus.
OC Archaea: Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A., SEQUENCE OF 150-184, AND CHARACTERIZATION.
RC STRAIN=VCI / DSM 3638 / ATCC 43587 / JCM 8422;
RX MEDLINE=96355370; PubMed=8702780;

RA Voorhorst W.G.B., Eggen R.I.L., Geerling A.C.M., Platteuw C.,
RA Stenzen R.J., de Vos W.M.;
RT "Isolation and characterization of the hyperthermostable serine
RT protease, pyrolysin, and its gene from the hyperthermophilic archaeon
RT Pyrococcus furiosus.";
RL J. Biol. Chem. 271:20426-20431(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VCI / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
RN [3]
RP CHARACTERIZATION, AND 3D-STRUCTURE MODELING.
RC MEDLINE=21079021; PubMed=11210516;
RA de Vos W.M., Voorhorst W.G.B., Dijkgraaf M., Kluskens L.D.,
RA Van der Oost J., Stenzen R.J.;
RT "Purification, characterization, and molecular modeling of pyrolysin
RT and other extracellular thermostable serine proteases from
RT hyperthermophilic microorganisms.";
RL Meth. Enzymol. 330:383-393(2001).
CC -1- FUNCTION: Has endopeptidase activity toward caseins, casein
CC fragments including alpha-S1-casein and synthetic peptides.
CC -1- SUBCELLULAR LOCATION: Cell-envelope associated.
CC -1- PTM: LHM pyrolysin seems to be produced by autoprolytic
CC activation of HWM pyrolysin.
CC -1- PTM: Glycosylated.
CC -1- MISCELLANEOUS: Thermostable; high activity at 95 degrees Celsius.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: U55835; AAB09761.1;
CC EMBL: AE010153; AAL80411.1;
CC HSSP: Q45670; IDBI.
DR MEROPS: S08.100; -;
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 4.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
KW Hydrolyase; Serine protease; Glycoprotein; Zymogen; Signal;
KW Complete proteome.
FT SIGNAL. 1 7
FT PROPEP 1 149
FT CHAIN 150 179
FT ACT_SITE 179 179
FT ACT_SITE 365 365
FT ACT_SITE 441 441
FT CARBOHYD 152 152
FT CARBOHYD 222 222
FT CARBOHYD 228 228
FT CARBOHYD 240 240
FT CARBOHYD 257 257
FT CARBOHYD 262 262
FT CARBOHYD 298 298
FT CARBOHYD 327 327
FT CARBOHYD 406 406
FT CARBOHYD 451 451
FT CARBOHYD 653 653
FT CARBOHYD 739 739
FT CARBOHYD 792 792
FT CARBOHYD 893 893
FT CARBOHYD 908 908
FT CARBOHYD 917 917
FT CARBOHYD 929 929
FT POTENTIAL.
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CHARGE RELAY SYSTEM (BY SIMILARITY).

FT CARBOHYD 1048 1048 N-LINKED (POTENTIAL).
 FT CARBOHYD 1056 1056 N-LINKED (POTENTIAL).
 FT CARBOHYD 1084 1084 N-LINKED (POTENTIAL).
 FT CARBOHYD 1117 1117 N-LINKED (POTENTIAL).
 FT CARBOHYD 1133 1133 N-LINKED (POTENTIAL).
 FT CARBOHYD 1140 1140 N-LINKED (POTENTIAL).
 FT CARBOHYD 1148 1148 N-LINKED (POTENTIAL).
 FT CARBOHYD 1208 1208 N-LINKED (POTENTIAL).
 FT CARBOHYD 1233 1233 N-LINKED (POTENTIAL).
 FT CARBOHYD 1237 1237 N-LINKED (POTENTIAL).
 FT CARBOHYD 1332 1332 N-LINKED (POTENTIAL).
 FT CONFLICT 607 609 AKA -> PKR (TN REF. 1).
 FT CONFLICT 881 881 Y -> H (TN REF. 1).
 SQ SEQUENCE 1398 AA; 154474 MW; 355D873A27D56552 CRC64;
 Query Match 3.1%; Score 138; DB 1; Length 1398;
 Best Local Similarity 19.5%; Pred. No. 1.4;
 Matches 157; Conservative 91; Mismatches 262; Indels 296; Gaps 41;
 185 PNCNHAATGCVATAAQAQIMRHSMPLOGESFDPYHAGSLVNSGTFEGEYDMINPGRP 244
 351 PNCGEYAVFG-----WDCHGHGT--HVAGTVAGYDSN--NDAMDWLSMYSGE 392
 245 -----DLNLNLTQ-----SQVDAVATLMDVSA----- 267
 393 WEVFSRLYGMWYTNVTTDVGVAPGAQIMAIRVLSNDGKSMMDIEGMTVAATHGADV 452
 268 VSMFS-----YENGSGTSYVV-----VG 286
 453 ISMSLGGNAPYLDGTDPESAVADELKEKYGVFEVIAAGNCPGINTVSGPVAATKATV 512
 287 ALBNRFRYKSLQLHVALYTSOEMHDMIRGELASGRPVY-----AGNOSIGHAFC 340
 513 A-----AAVYIN- GYVVSQ-----ALGYPDYGGYFPATYVRIAF-FSS 552
 341 DGVAADGTGFHFN-----GMGVSNGFYKLTLSPTSLGIGEGEGFTIYOEITIGIEPAKT 397
 553 RGRIGDEIKPYNVAAFGY-----YSSLPMTIG-----ADPMGTSMA-T 593
 398 PAEAGTDAIPIALKIDIEAEYS-----ESGLNAGSYLYTNGEESNLDLGR----- 445
 594 PHSVAVALLISGAKAGIYNDPIIKVLESCATMLEGDPYQOKTTELDOGHGLVNV 653
 446 -----LNKADGEYIEKTSINISWYCGHPSFSLAPNOLSGINTITLLRRIGTE- 499
 654 KSEIILKALINGTTLPIVDHMAKDSYSDFAEY-----LGVDVINGLALARNSPD 701
 500 --OMEPVRAHOGYVNSIKVNTDP-----NNVYTVDNNEKLSIVPNSFVADINSYE 551
 702 IYEMH- IKYVGTETEFRTFEIYATEPMIKPFVSGSVILENTEFVLRV-----KYDEGLE 755
 552 HSTITVQFNSDSDEIRTPVAFALSTGATADVISLGMVAEYVGGSSNYPVYWSKDVLT 611
 756 PGLYVGRITIIDP-----TTPV-----IIDEIL-----NTVIAPEK---F 787
 612 LSGDVTTL-WYRSINNOKDEMKKISVSYKTPETETHPLEFVGHNOTSYTTLDMANRY 670
 788 TPENNTLTWY-----DINGPEMYTHHEFTYPEGVDVLYAMTYDYG 830
 671 L--PD-----FTLKNL-----GLPENGELVYV-----FROGSS-----SSSLW 702
 831 LVRPDMFVFPYLODYLPAVAVSPMGMNEMELVMTGTFNAPLYSGELVRIYGVETPSW 890
 703 AAOET-----VHIKGETFVVKPVVEG--PIPDSSYRATLHAFNGOOLKGRNRY 753
 891 YIKRTLDNTERTSEIENITINIPATILIPIGLGTYNASVSV--GGEFFIKG-----944
 754 YKIVNGTAVEALIESSEELRVPNPARDVEYISACPIP-OETSIILFDLSGKIYMKNSLS 812
 945 -----IEVEGTAELK-----IRICNPSPVNSDDLVIYDSKGNLVALDGNP 986
 813 AGHGRMDVSRPLPGAVIILKVDGYTTR 838

DB 987 TAEVEVVEYKPKGVSYIVHGYSVR 1012
 RESULT 12
 CADN_HUMAN STANDARD; PRT; 3354 AA.
 AC Q9H251; Q9H4K9;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cadherin 23 precursor (Otocadherin).
 GN CDH23 OR KIAA1774.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, VARIANTS USHD M-1281 DEL;
 RP H-1436 AND Q-1746, AND VARIANTS C-3; A-490; N-436; T-1222; C-1349;
 RP N-1351; A-1575; S-1671; I-1675; Q-1804; S-1999; K-2044; Q-2358;
 RP L-2380; Q-2588 AND L-3125.
 RX MEDLINE-20578759; PubMed-11138009;
 RA Bolz H., Von Bredelrow B., Ramirez A., Bryda E.C., Kutsche K.,
 RA Nothwang H.G., Seeliger M., Del C., Salcedo Cabrera M.,
 RA Villa Caballero M., Pelaez Molina O., Gal A., Kubisch C.;
 RT "Mutation of CDH23, encoding a new member of the cadherin gene family,
 RT causes usher syndrome type 1D.";
 RL Nat. Genet. 27:108-112(2001).
 RN [2]
 RP SEQUENCE OF 803-3354 FROM N.A., ALTERNATIVE SPLICING, AND VARIANTS
 RP DNB12 N-990; N-2045; N-2202; N-2950; C-2956 AND T-3059.
 RX PubMed-11090341;
 RA Bork J.M., Peters L.M., Riazuddin S., Bernstein S.T., Ahmed Z.M.,
 RA Ness S.L., Polomeno R., Ramesh A., Schloss M., Srisailapathy C.R.S.,
 RA Wayne S., Belman S., Desmukh D., Ahmed Z., Khan S.N.,
 RA Klostian J.M.D., Li X.C., Lalwani A., Riazuddin S.,
 RA Blitner-Gindzicz M., Nance W.E., Liu X.-Z., Mistor G., Smith R.J.H.,
 RA Griffith A.J., Wilcox E.R., Friedman T.B., Korell R.J.;
 RT "usher syndrome 1D and nonsyndromic autosomal recessive deafness
 RT DNB12 are caused by allelic mutations of the novel cadherin-like
 RT gene CDH23.";
 RL Am. J. Hum. Genet. 68:26-37(2001).
 RN [3]
 RP SEQUENCE OF 2173-3354 FROM N.A. (ISOFORM 4).
 RC TISSUE=Brain;
 RX MEDLINE-21481446; PubMed-11597768;
 RA Nakajima D., Nakayama M., Kikuno R., Hirose M., Nagase T., Ohara O.;
 RT "Identification of three novel non-classical cadherin genes through
 RT comprehensive analysis of large cDNAs.";
 RL Brain Res. Mol. Brain Res. 94:85-95(2001).
 CC -1- FUNCTION: Cadherins are calcium dependent cell adhesion proteins.
 CC They preferentially interact with themselves in a homophilic
 CC manner in connecting cells. Cadherin 23 may function as hair
 CC bundle organizer perhaps by cross-linking the stereocilia (by
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (by similarity).
 CC -1- ALTERNATIVE PRODUCTS: At least 4 isoforms: 1 (shown here), 2, 3
 CC and 4; are produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Particularly strong expression in the retina.
 CC Found also in the cochlea.
 CC -1- DISEASE: Defects in CDH23 are the cause of usher syndrome type 1D
 CC (USHD). Usher syndrome type 1 is an autosomal recessive disorder
 CC characterized by congenital profound sensorineural hearing loss,
 CC constant vestibular dysfunction and prepubertal onset of retinitis
 CC pigmentosa leading to blindness. Usher syndrome is the most common
 CC cause of combined deafness and blindness in developed countries.
 CC -1- DISEASE: Defects in CDH23 are the cause of a form of nonsyndromic
 CC autosomal recessive deafness (DFNB12).
 CC -1- SIMILARITY: Contains 27 cadherin domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.ch).

[illegible][illegible]

```

Db 452 NENDNRPISQPLYNISLSENVTVGTSVLTATDNDAGTGFGEVSEFSDPDRFSL--- 508
Qy 177 MNKKPELPLPGNNAIYGCVAATA-----AQIMRHSNPLGEGSEFDYHAGLVGNMGSFTG 232
Db 509 --DKD-----TGLIMLIARLDYELLQFTLLIARDGGGEETTGVRN----- 550
Qy 233 EMDWIMPGNPDLNLTOSVDAYATLMRDVSASVM-----SFENGSGTYSYV 283
Db 551 -VLD-VN-----DNVPTFOKDAYVGALENPEPSTQVLRLATDEDSPPNNOITYSIY 601
Qy 284 VVGLANNNERYKRSLOLHVRLATYSOEHMDIRGELASGRPVYAGNNOISIGHAFVODGY 343
Db 602 SASAFSGSYFDI-----SLYEGY-----GVATSVRPLDYE----- 630
Qy 344 AASGTHFFNMNGMGCVNGNPFKLLLSPTSLGIGEGGFIYIYQETITIGEPARTPA----- 399
Db 631 -----QISNGLIYLTVM-----AMDAGNPPLNSTVPVTIEVFENDNPPTFSKP 674
Qy 400 -----EAGTDALPILALK-DIEAEYKSSGSLNGVSIYNTGEGOSNLDLGYRLNK 448
Db 675 AIFVSVVENIMAGATYVLFILNATDLDSREYGOES-----IISL--EGSTO-----FRINA 723
Qy 449 ADGEVY-----EYKTSINISWY--GYGEHPESFSLAPNQLSOGINT--ITLLYRT 496
Db 724 RSGEITTTSLDRETKSEYILYIYRAVDGCVGHNOKT-----GIATVNTITLDDIND 773
Qy 497 GTEGOWEVRHAGGYNSIKVNTDPN--NVYVTVDDNECK-----LSIYNSVADLNSY 550
Db 774 NHPTMDAPY-----YINLVEMTPDSDVTVAVADDLGNGTLVYSIOPNKFYSLNST 829
Qy 551 --EHSTITVOFNSDSD-----EIRTPVAFALSTGATADVISL----- 587
Db 830 TGRIRTHAMLDRENDPHEAELMRKIVSVTDCGRPLKATGSATVFAVVLDDNNDPT 889
Qy 588 -----GWMAEVPGGSSNYPYVNSKDVLLTSEGYTTLTMRISINNQDEMKKISVS 640
Db 890 FQNLPRVAEYLEGIAPAGVSIYOVY-----AIDDEG-----LN-----GLVSY 927
Qy 641 KPTETHTPLFEVGHNOTSTYTLDMAHNRVLPDTLK-----NLGLPENGELVYVFPQOTS 966
Db 928 RMVGNPRMDFLINSSGVYVTTTELDREIATYQLRVASDAGTP-----TIS 976
Qy 697 SSGSLMAADEVYHI-----KOGETF--VYKPVVEGRIP-----DGSYRATLH 736
Db 977 SSTSL-----TIHVLVDNDETPTEFPYAVNVSVSEDPREFRYVMJLCTDNDVGLNAELS 1031
Qy 737 AIFVNG--QOOLYLKGRNTYKLVNG-----TA-----VEATSSSEIRVFPNPADYV- 783
Db 1032 YFTTGGVNDGKESVGYRDAVVRVVGDLDBETTAAYMLILEALDNGFVGRKHTGTATVFTY 1091
Qy 784 -----EISAP-CIPOETSTILFDLS-----GKI--VMKNSLSAGHG--R 817
Db 1092 VLDVNDNRRPIFLOSSTEASVPEDIPGCHSILQKATDADEGEGRWYRILGNHGNR 1151
Qy 818 MDVS-----RLPNGAYILKVDGY 835
Db 1152 IHVSNGLMRGRPLDRERNSSHVLIVEAY 1181

```

RESULT 13

```

ID GUNZ_ERWCH STANDARD; PRT; 426 AA.
AC P07103;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endoglucanase Z precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase Z)
DE (Cellulase Z) (EGZ).
GN CELZ OR CEL5.
OS Erwinia chrysanthemi.
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae;
OC Pectobacterium.

```

```

OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3937;
RX MEDLINE=8216177; PubMed=2835589;
RA Gidasppl A., Caml B., Aymeric J.-L., Ball G., Creuzet N.;
RT "Homology between endoglucanase Z of Erwinia chrysanthemi and
RL endoglucanase of Bacillus subtilis and alkalophilic Bacillus.";
RN Mol. Microbiol. 2:159-164(1988).
RN [2]
RP REVISIONS, AND DISULFIDE BOND.
RC STRAIN=3937;
RX MEDLINE=94203057; PubMed=8152378;
RA Bortoli-German I., Brun E., Py B., Chippaux M., Barras F.;
RT "Periplasmic disulphide bond formation is essential for cellulase
RL secretion by the plant pathogen Erwinia chrysanthemi.";
RN Mol. Microbiol. 11:545-553(1994).
RN [3]
RP MUTAGENESIS, AND DOMAINS.
RX MEDLINE=91312880; PubMed=1677466;
RA Py B., Bortoli-German I., Hachez J., Chippaux M., Barras F.;
RT "Cellulase EGZ of Erwinia chrysanthemi: Structural organization and
RL importance of His98 and Glu133 residues for catalysis.";
RN Protein Eng. 4:325-333(1991).
RN [4]
RP STEREOCHEMISTRY OF THE REACTION.
RX MEDLINE=92225124; PubMed=1563515;
RA Barras F., Bortoli-German I., Bauman M., Rouvier J., Gey C.,
RL Heyraud A., Henrissat B.;
RT "Stereochemistry of the hydrolysis reaction catalyzed by
RL endoglucanase Z from Erwinia chrysanthemi.";
RN FEBS Lett. 300:145-148(1992).
RN [5]
RP STRUCTURE BY NMR OF 365-426.
RX MEDLINE=98070232; PubMed=9405041;
RA Brun E., Morlaud F., Gans P., Blackledge M.J., Marion D.;
RT "Solution structure of the cellulose-binding domain of the
RL endoglucanase Z secreted by Erwinia chrysanthemi.";
RN Biochemistry 36:16074-16086(1997).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.30 ANGSTROMS) OF 44-335.
RX MEDLINE=21392910; PubMed=11501995;
RA Chapon V., Czjzek M., El Hassouni M., Py B., Juy M., Barras F.;
RT "Type II protein secretion in Gram-negative pathogenic bacteria: the
RL study of the structure/secretion relationships of the cellulase cels
RL (formerly EGZ) from Erwinia chrysanthemi.";
RN J. Mol. Biol. 310:1055-1066(2001).
RN [7]
RP FUNCTION: REPRESENTS 97% OF THE GLOBAL CELLULASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: Y00540; CAA68604.1; -
PIR: S03767; S03767
PDB: 1ATW; 06-MAR-98.
PDB: 1EGZ; 26-MAR-99.
InterPro: IPR003610; CBM_5_12.
InterPro: IPR001547; GH_5.
Pfam: PF00150; cellulase; 1.
Pfam: PF02839; CBM_5_12; 1.
SMART: SM00495; ChIB3; 1.
PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.
KW

```



```

Db      586 -----LEITSANSECTAVL-----FGGDYFKKO-----YNT-----612
QY      436 EGSNLDLGYLNKADGEVIEVKYKSSINISW-----YGEHPESESAPNQ 481
Db      613 ---SNLN-----AINNTNTILNMSNDSATLVFNVLGNYSYSEBDNLLA-----652
QY      482 LSGGINTYLLYRRGTGEQNEPRHAGCGVNSIKVATDTPNNV-----VYVVDNNEGKLSI 558
Db      653 -KGFAPKILEFNNGTNTNISIKGVASGSISITDGTGETIENIMLENTVFNKDAKSYSF 711
QY      539 -VNSFVADINSTYESTTVOFNS---DSPDEIRTPVAFALSTGATADVISLGWMAE 593
Db      712 NLNLNIMAVNKSAYELYNPFPKSIWIDGSNTTIP-----NID 751
QY      594 VPGGS-----SNYPVMSKCVLTLSDDYTLNTRFSINOKDEKKTIGSVK 642
Db      752 IIPGEVWNSKTYFTESGVPITWANCFSITLSKRDYIL-----LNEVSQIGSSYV-- 800
QY      643 PREYTHPEVEGHNGTSTYTLDMANRVLDP-----FTLKNLGLPENGELVYVF-- 691
Db      801 ---VVEIYVVG---SYLIKT-KHIYPDAGCTDIYIIVENISVKTPEIYVYDLI 851
QY      692 --KQTSSSGSLMAOETHIKOG-ETEVYKP-----VVEGPIPGSYRATLH 736
Db      852 PKNFTYVSD--VWVNSMLIAEGNTTITNPYYNLMMWALHAIYPGADGNWMDFAE 908
QY      737 AEVNGQOOLYTLGKRNYTVKINGTAVEAIESESEIRVPPNAPADVVELSAPCI 790
Db      909 ILAKTVVLIHYK-----LNGIG-EPYSDAFTVGIDPINSILPTTSPKI 951

RESULT 16
ID      AMY_BUTFI      STANDARD;      PRT;      976 AA.
AC      P30269;
DT      01-APR-1993 (Rel. 25, Created)
DT      01-APR-1993 (Rel. 25, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Alpha-amylase precursor (EC 3.2.1.1) (1.4-alpha-D-glucan
DE      glucanohydrolase).
DE      GN
OS      Butyri vibrio fibrisolvens.
OC      Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae;
OC      Butyri vibrio.
OX      NCBI_TaxID=831;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-H17C;
RX      MEDLINE=91286207; PubMed=2061294;
RA      Rumbak E., Rawlings D.E., Lindsey G.G., Woods D.R.;
RT      "Cloning, nucleotide sequence, and enzymatic characterization of an
RT      alpha-amylase from the ruminal bacterium Butyri vibrio fibrisolvens
RT      H17C."
RL      J. Bacteriol. 173:4203-4211(1991).
CC      -1 CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic
CC      linkages in oligosaccharides and polysaccharides.
CC      -1 SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC      KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: M62507, AAA22005.1, -
DR      PTR: A42466; A42466.
DR      HSSP: P00691; IBAG.
DR      InterPro: IPR000461; Alpha_amylase.
DR      InterPro: IPR000585; CBM 25.

```


DR Pfam: PF00128; alpha-amylase; 1.
 DR Pfam: PF02806; alpha-amylase_C; 1.
 DR Pfam: PF03423; CBM25; 1.
 KM HydroLase; Glycosidase; Carbohydrate metabolism; Signal.
 FT SIGNAL 1 33
 FT CHAIN 34 976
 FT ACT_SITE 323 323
 FT ACT_SITE 327 327
 FT ACT_SITE 447 447
 SQ SEQUENCE 976 AA; 106695 MW; 61D690EDC19315AC CRC64;

Query Match 3.0%; Score 134.5; DB 1; Length 976;
 Best Local Similarity 19.4%; Pred. No. 1.4;
 Matches 155; Conservative 116; Mismatches 259; Indels 267; Gaps 47;

QY 13 FGLA-----HOGSAPYTKERALSRLR-----ALROYSLRMGTAVSDKISIDYVYRQG 62
 DB 232 YGVAIVIDLPNHTTPTSTGSLAKALMEAGGSDALYHATGKIG-GGYTDRLRLTY-YSMG 289
 63 DAERGITSOESESPAYFYVANRG-----NNEGVALVA-----DDRPTILAYSPIGRFD 112
 290 GLP-DVTEENTGFOQYFTEFLKDCVYLGADGFRIDTAKHISLPDDPPVS--DYSAGR-- 344
 QY 113 MDSMPDLRMWLOIYDDEIGLISGAQLNEELRTGVPALVMDNGHFADPMKRN 172
 DB 345 -NFEYPMNRREALNEXSEVEGKSYDELFEVYGEVLGTNDRLAAVQOYIGGTAS-----N 398
 QY 173 QGYVMNKKELPLPNCNATYGCVATAAQIMRHSWPLQSGSPDYHAGSLVGMWSGTEG 232
 DB 399 YG---SSLRSLSSGN-----LSVNRLLDYOIYDDTAAGS-----TYTADTE 437
 QY 233 EYMDIMNPGN--PDLNLTQSOVDAYATLM-----RDSASVSMFYENGSG--GYTS 281
 DB 438 KLYTWVSHDNYMNDSESCKS-IDDDMVIMGSIITAKRGCTLFTRSPNNSAENPYG 496
 QY 282 VYVVGALR-----NMFYK-----RSQQLHYRALYTSQEHDM 314
 DB 497 DNLIGAAGSPIYKAPEYKAVLFRKKGEADEYLSPNGNIGTLM-----ERYNDT 548
 QY 315 IRC-----ELASG-RPYVYAGNNOISGAFVCDGAS----- 345
 DB 549 VQGVAVIYMAOTRTTISTETHLSGDIYPPQVEGNSVYF--LVKDGVLSSGVEGQVYVL 605
 QY 346 ---DGT---FHFNMGCGV-----SNGFYKLTLLSPSLGIGRGI 380
 DB 606 SEKMDGTGKYVSFFNNKNNMGVAVRVDAEETLDTIDENCGWFOYTYLDE----- 656
 QY 381 GFTIYOEITIGIEPA---KTPAEAGTDALPILALKDIEAEY---KSEGLNV-GYSIY-- 431
 DB 657 -FTIRFESADCKEVSPEFQITAEAGTFATP---DSSELYSKAEAEGLGITHYYPYFF 711
 QY 432 ---NTGEQSN--LDLGRYLKADGEVIEVKTSSINISWYGGHSPESFSLAPMQLSOGI 486
 DB 712 NTEWNGSVYTYGMLDGAQLFGGMPGVAVANEGS--GWY-----RADYKTGCEI 758
 QY 487 NTIRLLRRTGTEQMEVBRHAGGVNSIKVN--TTDPNNVYTVVONNEGSLTVIPNSFV 544
 DB 759 TAPULIIN-----NGNGIOTVNEGITEIPDSKDIYLLDAEKSGOQLIVNY- 804
 QY 545 ADLINSYEHSS-----TITVOFNSDPELIRTPAFALSTGATA-----DDVI 585
 DB 805 EDRKSAKALGVSGSYTAYFYNTGMDKV-----CAWTGATATLGDMPGKELTQDD- 857
 QY 586 SLGVMAAEVPEG--SSNPVYVNSKDVLTLSGDYT-----LWTFRSINN-----Q 628
 DB 858 --GYSVYVLPAGBEDDNIITFNNG-----NNCKQTNDMKISDMKYRFLNNGISYQKXGSK 911
 QY 629 KDEWKTI---GSYSVKI 642
 DB 912 KDAMEALAGAGDYET 928

RESULT 17

HEC1_HAEIN
 ID HEC1_HAEIN STANDARD: PRT: 837 AA.
 AC P33397;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Outer membrane usher protein hifc precursor.
 GN HIFC.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 CC Haemophilus.
 OX NCBI_TaxID:727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bagan / Serotype B;
 RX MEDLINE=94131580; PubMed=7905461;
 RA Watson W.J., Gilstorf J.R., Tucci M.A., McCrea K.W., Forney L.J.,
 RA Mairs C.F.;
 RT Identification of a gene essential for pilation in Haemophilus
 RT Influenzae type b with homology to the plus assembly platform genes
 RT of Gram-negative bacteria.";
 RL Infect. Immun. 62:468-473(1994).
 CC -1- FUNCTION: ESSENTIAL FOR PILATION.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
 CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-97 IS THE INITIATOR.
 CC
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC
 CC EMBL: U02932; AAB53096.1;
 DR InterPro: IPR000015; FimB_usher.
 DR Pfam: PF00577; Usher; 1.
 DR PROSITE: PS01151; FIMBRIAL_USHER; 1.
 KW Outer membrane; Transmembrane; Fimbria; Transport; Signal.
 FT SIGNAL 1 26
 FT CHAIN 27 837
 FT DISULFID 813 833
 SQ SEQUENCE 837 AA; 92733 MW; BA529323AF63BF8B CRC64;

Query Match 3.0%; Score 133; DB 1; Length 837;
 Best Local Similarity 21.3%; Pred. No. 1.4;
 Matches 129; Conservative 80; Mismatches 212; Indels 186; Gaps 32;

QY 22 APYTKERALSRLR-----QVSLRMGTAVSDKISIDYVYRQGDAREGITSQDEG 74
 DB 282 APVYGVANINAKSISIKONGYTIYQIVPAPFVINDIYASGY---SGDLTVEI-QESDG 337
 QY 75 SPAYFYVANRGNNGGYLVADDRIPITLAVSPGPFMDMPNLRMWLOIYOET--G 132
 DB 338 KYRSFYIP-----FSNLAPLMRYGHLRYLAGGRIRIDS-----RTFERVYLOG 381
 QY 133 LI---LSGKAQLNEELRTGVPALVMDNGHFADPM--RMNGYPMNKK--PLPNG 187
 DB 382 VLOGYGLNHLNLSLTYRHYRAGLF-----GGLMTPIGAFAADATWSAEPFLKYS 436
 QY 188 NHAYT-----GCVATAAQIMRHSWPL----- 210
 DB 437 KNGYSLGSYSINFEINIGTNLTLLAA--YRYSRDFYTLSDYIGLNRTRFGSGAYLPEIY 494
 QY 211 QGBSFPYHAGSLVGNMNGTF--GEYMDIMNPGNPDLNLTOSQVDAYATLMVDYASV 268
 DB 495 RPKNOFOVSLQSGISGNNGNLTLSGQTYNWEKRTN-----TOYOV-AYSSEHILINYSV 548
 QY 269 SMS---EYENGSGTYSVYVVGAL-----RNMFYKRS-----QLHYRALYTSQEHDM 314
 DB 549 NLSQSIDKFKCKRNSITLSLPLDGNHSDSSYSRSGNDINRGLVNGSF--GERHOW 606

OY 315 IREGLASGRVYVYAGNNSIGHAFVCDYAS-DGTFHFMNGCVSNGEYKLTLS-PTS 372
 DB 607 SYGINS-----RNNQ-----GRTSYDGNLSHNSISYASYSRDLKRNSTIS 650
 OY 373 LGTGGEGT---GFTYQEIITIGIEPAKTPAEAGTALPLALKDIEAEYKSSGLNVGY 428
 DB 651 LGVSGAVVAHKYGITLSQPV-----GSEFALITHAKD-AAGAKVESGANVSL 695
 OY 429 STYNTGEOSNDLGLRKNKADGEVIEKTSINISWYG-----GEHPESFLAPNLSQ 484
 DB 696 DYF-----GNAAVPTSPYEIVYIGINPSDAANVEFEATERQILP 736
 OY 485 GINTITLYRRTGTEQMEPRHQAQGYNSIKYNTDPNNVV-----TYDNNEGKLSIYP 540
 DB 737 RANISILVDERTCKN-----TWLFNLPLNPEEPVPASTADDSG----- 777
 541 NSFVADL 547
 778 -AFVGDV 783
 RESULT 18
 EAE_ECO57 STANDARD; PRT; 934 AA.
 AC P43261; 085627;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Intimin (Attaching and effecting protein) (Eae protein) (Gamma-
 DE Intimin)
 GN EAE OR EAEA OR 25110 OR ECS4559 OR L0025.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=83334;
 RN 11)
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RA MEDLINE=92204008; PubMed=1552854;
 RX Yu J., Kaper J.B.;
 RT "Cloning and characterization of the eae gene of enterohaemorrhagic
 RT Escherichia coli O157:H7".
 RL Mol. Microbiol. 6:411-417(1992).
 RN 12)
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / CL-8;
 RA MEDLINE=92249750; PubMed=1577255;
 RX Beebekhee G., Louie M., de Azavedo J., Brunton J.;
 RT Cloning and nucleotide sequence of the eae gene homologue from
 RT enterohaemorrhagic Escherichia coli serotype O157:H7.";
 RL FEMS Microbiol. Lett. 70:63-68(1992).
 RN 13)
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RA MEDLINE=98339885; PubMed=9673266;
 RX Perna N.T., Mayhew G.F., Postal G., Elliott S., Donnenberg M.S.,
 RA Kaper J.B., Blatterner F.R.;
 RT "Molecular evolution of a pathogenicity island from enterohaemorrhagic
 RT Escherichia coli O157:H7".
 RL Infect. Immun. 66:3810-3817(1998).
 RN 14)
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / DEC 3a, and O157:H- / DEC 3f;
 RA MEDLINE=99261633; PubMed=10331248;
 RX McGraw E.A., Li J., Selander R.K., Whittam T.S.;
 RT "Molecular evolution and mosaic structure of alpha, beta, and gamma
 RT Intimin of pathogenic Escherichia coli".
 RL Mol. Biol. Evol. 16:12-22(1999).
 RN 15)
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dinalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blatterner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7".
 RL Nature 409:529-533(2001).
 RN 16)
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / RMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shingawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN 17)
 RP SEQUENCE OF 1-37 FROM N.A.
 RC STRAIN-O157:H7 / HAI;
 RX MEDLINE=96079490; PubMed=8566710;
 RA Zhao S., Mitchell S.E., Meng J., Doyle M.P., Kresovich S.;
 RT "Cloning and nucleotide sequence of a gene upstream of the eae gene
 RT of enterohaemorrhagic Escherichia coli O157:H7".
 RL FEMS Microbiol. Lett. 133:35-39(1995).
 CC -1- FUNCTION: NECESSARY FOR THE PRODUCTION OF ATTACHING AND EFFACING
 CC LESIONS ON TISSUE CULTURE CELLS.
 CC -1- SUBCELLULAR LOCATION: OUTER SURFACE.
 CC -1- SIMILARITY: BELONGS TO THE INTIMIN/INVASIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 LYSM REPEAT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; Z11541; CAAT7642.1; -;
 DR EMBL; X60439; CAA42967.1; -;
 DR EMBL; AF071034; AAC31504.1; -;
 DR EMBL; AF081182; AAD05499.1; -;
 DR EMBL; AF081183; AAD05499.1; -;
 DR EMBL; AF000595; AAC58823.1; -;
 DR EMBL; AF000256; BAB37982.1; -;
 DR EMBL; U32312; AAB0011.1; -;
 DR InterPro: IPR003344; B1g_1.
 DR InterPro: IPR003343; B1g_2.
 DR InterPro: IPR003535; Intimin.
 DR InterPro: IPR002482; LySM.
 DR Pfam: PF01476; LySM; 1.
 DR Pfam: PF02368; B1g_2; 1.
 DR Pfam: PF02369; B1g_1; 2.
 DR PRINTS: PR01369; INTIMIN.
 DR SMART: SM00257; LYSM; 1.
 KW outer membrane; Virulence; Complete proteome.
 FT REPEAT 65 113
 FT CONFLICT 221 221
 FT CONFLICT 311 312
 FT CONFLICT 318 318
 FT CONFLICT 642 642
 FT CONFLICT 769 769
 FT CONFLICT 771 772
 SQ SEQUENCE 934 AA; 101835 MW; BE73686D/F79D3C9 CRC64;
 Query Match 3.0%; Score 133; DB 1; Length 934;
 Best Local Similarity 17.2%; Pred. No. 1.6;
 Matches 165; Conservative 128; Mismatches 338; Indels 326; Gaps 44;
 OY 74 GSPAYFYVANR--GNNEGVALVAADRIPITLAYSPIGRF-----DMSMPD-N 119

```

Db      26 GIGLFFVYVQNSPANEENFKLGSDSKILLHDSYQ--NRLFYTLKTGETVADLSKSDIN 83
QY      120 L-RM---LQIYQELGLISGKAQNLIELRTGEPVAEVHALMDNCHFANDPRKRNQGY 175
Db      84 LSTIWSLNRHLVSESEEMKAAIPGQ--QITLPLKRLPEFSAL----- 124
QY      176 PMNKEPPLPNC-----NHAUYGVATAAQAQIMRYHSPLOG 212
Db      125 PILGSAPIVAGVAGHTNKLTKMSPDVTKSNMTDDKALYYAAQQAASLSQSLSRLNG 164
QY      213 EGSEFDYHAGSLVGNMSTGEGEMDWIMNMGNDPDLNLTQSQVADATLMDVASVSMSE 272
Db      185 DYAKDPAALG-IAGNOAASS-QLOAWLOHYGTAEVNLQSGNNPFG--SSIDFLPPE 234
QY      273 YEN-----GSGT-----YSYV-----VGL 268
Db      235 YDEKMLAFCQVAGARIIDSRFTANLGAQGFLLPANMLGYNVETLDDSGDNTRLGIGGE 294
QY      289 RNNFRYKRSLOLHVRLALYTSQEHMDMIRGELASGRPY-----YAGNNOSIGHAFVC 340
QY      295 YMDYFRKS---VNGYFRMSGWHEYNKRDYDERPANGPDIFNGILPSPYALGALYI 350
QY      341 DGASDGTFRHNGMGVSNGLFYKLLLSFTSLGIGGEGITTYOEITIGIEPAKTPAE 400
Db      351 EQYYGDVNAFLNSD-----KL-----QSNPGAATVGVNTPIPILVTMGI----- 369
QY      401 AGTDALPILAKLIEAEKSESLANGYSTYNGEES---NIDLGRILKAKAGEVIEVK 457
Db      390 -----DYS-----HGTGENEDLLYSQFIRQFQKSMQOILEPQ 422
QY      458 TSSINISWYGYGHEPESFLAPNQLSOGITITLLYLR-----TGTEQ----- 500
Db      423 -----YVNEKLTSSRYDLVQRNNNITILEYKKODIISLNPHDNGEHSQKI 472
QY      501 -----W-EVPRHAOGGYVNSIKVNTDPNNVNVTVDNNEKLSIVPNSFVA 545
Db      473 OLIVKSKYGLDIRIYMDSDALSRSOGGOIQHSGSQAODYQALIPAYVQGGG-----NITYV 527
QY      546 DLSYSHS-----TIVQFNSDSPDEIRTPVAFALSTGATADVISLGVMAVPG 596
Db      528 TARAIRNGNSSNNVOLTITVLSNGOVVQVDTDFADTKTSKADADITITTAIVKRN 587
QY      597 G---SSNYPVYMSKDVLLSEGDYTLWYRFISNNOKDEMKIGSVYTPREYTHLPREVG 654
Db      588 GVAQNAVPSFN-----IYSGTATL--GANSKAKIDANKATVTLKSSIP-----G 630
QY      655 HNOTSTYTLDMAHNRVLPDFTLKNLGLFENGELVVVFROTQSSSGSLMAAQETVHIKQGE 714
Db      631 QVVVASAKTAEMTS-----ALNASAVIFPDQTKASITEL-KADKTTAVANGK 675
QY      715 -----TFYKKPVYEG-PIPDGSR-ATLHAFVNGGOQLYL-----KGRNTY 753
Db      676 DAIKYVYVWKMNQOPVNNOSVTPSTNFGMFGKSSQTOATTDGNGRATITLSSAGKATV 755
QY      754 TVKIVNGTAVEAIESS--EIRV-----FPPARDYVEISAPC----- 789
Db      736 SATVSGAEKATKAEVFFDELKIDNKVDIIGNVRGELPRIMWQYGFKLKASGGDCTYS 795
QY      790 -IPQESIIIFDLGKIYVKNLSLASHGRMDVSRLPNG-----AYILKVGYYTKIN 840
Db      796 WSESTNISIAIVDASGRVTLN-----GKGSV-VIKATISGDQKOTVSYTIKASYMILKV 846

```

```

RESULT 19
CSD2_ECOLI
ID CSD2_ECOLI STANDARD; PRT; 802 AA.
AC P53513;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Outer membrane usher protein csd2 precursor (Csd6 fimbria usher protein).

```

```

GN CSD2.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=88775;
RA Wolf M.K., de Haan L.A.M., Cassels F.C., Willshaw G.A.,
RL Gaestel E.C.M., Gaastra W., Warren R., Boedeker E.C.;
SUBMITTED (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF C6 FIMBRIAL
CC SUBUNITS ACROSS THE OUTER MEMBRANE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC (By similarity).
CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: U04846; AAB51364.1; -.
DR InterPro: IPR000015; Fimb_usher.
DR Pfam: PF00577; Usher, 1.
DR PROSITE: PS01151; FIMBRIAL_USHER; FALSE_NEG.
DR KMW Outer membrane; Transmembrane; Fimbria; Transport; Signal.
FT SIGNAL 1
FT CHAIN ? 802 OUTER MEMBRANE USHER PROTEIN CSD2.
SQ SEQUENCE 802 AA; 90393 MW; 5BFDOB617F5B580 CRC64;

Query Match 3.08; Score 132.5; DB 1; Length 802;
Best local similarity 18.88; Pred. No. 1.4;
Matches 147; Conservative 102; Mismatches 275; Indels 259; Gaps 42;

133 LIISGKAQLNEEI-LRTGEPVAEVHALMDNCHFANDPRKRNQGYPMNKEPPLPNCNHA 190
132 LFTAVPKNLSEIDKISKDIDNGIALFFV-----YQVNTK--LANKNR 175
QY 191 YFGCVATAAQAQIMRYHSPLOGEGSEFDYHAGSLVGNMSTGTF-----GEM 234
Db 176 YD---YISVPNINPYFMRRL--NLFEFQNNDEKTEBRYNTYLEKSFYDKILNLVYGES 230
QY 235 YDWINMFGN-----PDLNLTQSQVADATLMDVASVSMSEYENG----- 276
Db 231 YTNNSVYNNYSFTGISYSTDTDMYTPSEID-YTPEIIGVADSDQIIVROGNTIITINESV 289
QY 277 -SGTYSYVVVGAALRNRRYK-RSLQLHVRALYTSQEHMDIRGEL--ASGRPY----- 326
Db 290 PAGEFSPT-----TNLMYGGQLNVAITDIYGNKKOYTVNNSSLPVARRAGLVNVPIS 344
QY 327 --YAGNNOSIGHAFVCDGYASDGTGFHFN---WCGVGVSNGLFYKLLSLPTSLGIGEGIG 381
Db 345 GKLLKKKSEGD-FTFOGDLNIGYR-HYNSLFLGQYQSKYFNLSITGIDLGSGAML- 401
QY 382 FTIYQELITGIEPAKTPAEAGTDLPLALKLIDAEYKSESGLNVGS-----IYVGE 435
Db 402 -----LHVSRSNKNNGYININQOQNTQLRPFVAG- 431
QY 436 EQSNLDLGYRLNKADGVEIVKTSINISWYG--YGEHPESFLAPNQLSOGIN---TIT 490
Db 432 --VNEFDIAYKKR---YVELS---DIGWGNLYNLQKNSFSL--LSKSLKYNFIS 478
QY 491 LLYRRTGEQMEPEVRAHAGGYVNSIKVNTDPNNVNVTVDNNEKLSIVPNSFVADLNS 550
Db 479 LDVNRK--MKYWD-----NAQDSNMSIRYFFKFRAMNITTNCSLNTKKQSY 521
QY 551 EHSITIVQFNSDP---DEIRTPVAFALSTGATADVIS-----LGM----- 589
Db 522 EKKDKRFSINISLPLTKDYGHISNYSFSNANGTATSSVGLNSSFNDARLMMNIQNR 581

```

QY 590 -----VMAEVPDSSNYP-----VWMSKD--VLTLSRG 615
 Db 582 TTRNNGYTNDNTSYATISAPYGVFTGSGSKYSQSFYSAGSGIVLHSDGAFQKAG 641
 QY 616 DYTLMATRESINNOCKEMKIGSVKTPREYR--HPLFEVGHQNTSTYTLDMANRYLDP 673
 Db 642 DSALVR--IDNTSD--IKIGN--TPGVYTGNGFALIPHLQPFKKWTILLINDKIDP 693
 QY 674 -FLKMLG--LPFNGELV-----VVFROGSSGSLMAADEVTHIKGFTFY 718
 Db 694 GTTLANIKKOVIPSRKAIVKVKDAAKNDILFKLT-TYDGKTPPLGAIHAERNG----- 747
 QY 719 KPYVEGPIPDGSYRATLHAFAVNGQOLYLKGRKNTYKIVNGTAVEATSESEIRVFPMP 778
 Db 748 KQNTGIVDDG-----MLYMSG-----LSGTGIINTVMNGKVCSPFS 786
 779 ARD 781
 787 END 789

RESULT 20
 VG37_BPARI STANDARD; PRT; 1103 AA.
 AC 09G0B5;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Long tail fiber protein p37 (Protein Gp37) (Receptor recognizing
 protein).
 GN 37.
 OS Bacteriophage ARI.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
 OC T4-like viruses.
 OX NCBI_TaxId=66711;
 RN 111
 RP MEDLINE=2048545; PubMed=11029414;
 RA Yu S.L., Ko K.L., Chen C.S., Chang Y.C., Syu W.J.;
 RT Characterization of the distal tail fiber locus and determination of
 RT the receptor for phage ARI, which specifically infects Escherichia
 RT coli O157:H7.
 RL J. Bacteriol. 182:5962-5968(2000).
 CC -1- FUNCTION: STRUCTURAL COMPONENT OF THE DISTAL-HALF TAIL FIBER. IT
 CC CONSTITUTES THE PART OF THE LONG TAIL FIBERS THAT RECOGNIZES THE
 CC BACTERIAL RECEPTOR (BY SIMILARITY).
 CC -1- SUBUNIT: THE DISTAL HALF-FIBER CONTAINS TWO MOLECULES EACH OF GP36
 CC AND GP37 AND ONE MOLECULE OF GP35 (BY SIMILARITY).
 CC -1- MISCELLANEOUS: THE TWO GP37 PROTEIN CHAINS RUN IN PARALLEL, THE
 CC LENGTH OF THE DISTAL HALF-FIBER, WITH THE AMINO END NEAR THE
 CC CENTER KINK OF THE FIBER AND THE CARBOXYL END AT THE DISTAL TIP.
 CC THE OTHER POLYPEPTIDES ARE DISTRIBUTED UNIFORMLY ALONG THE LENGTH
 CC OF THE DISTAL HALF-FIBER.
 CC -1- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: AF208841; AAG29754.1;
 KW Fiber protein.
 SQ SEQUENCE 1103 AA; 119177 MW; D7496C81C1DB056E CRC64;
 QY Query Match 3.0%; Score 132.5; DB 1; Length 1103;
 Best Local Similarity 19.8%; Pred. No. 2.2; Mismatches 235; Gaps 36;
 Matches 142; Conservative 84; Indels 255; Indels 235; Gaps 36;
 171 WNOGYPMNKEPLPNGNHAYTGVAT-----AAQIMRYHSWPLQEGSFYHAGSLVG 225

Db 68 YNOTGDTYLTNGFTQTGNFNLGIAVRTDIIAAGIMT-----EG-----GELIT 113
 QY 226 MNSGT-----EGEAYD-----WINPGNPDLNLTQSOYDAVATLMRPVSAVSMSFEYENS 277
 Db 114 KSGGTAAVHFPHSDADEKRIITSPAN--DGLTQVYN--IKRVYKASSESTFAHGN 167
 QY 278 GYSYVYVVALRNNEFYKRSLOLHVRYALYTSQEMHDMINGELASGRPVYAGNN----- 331
 Db 168 GLESPPEV-----FGWK--SVSTPVITNKV--ITKKYKDDVDIISMDNVPLSRI 215
 QY 332 -OSIGHAYVCDGYASDGTN-----PNNCGW-----GVS 359
 Db 216 TTAINHLYRVNRNAVSGIFHEVYKNDGITWYSGDIDAYLWSTWGGIKSSHISISIGLT 275
 QY 360 NGFYKLTLLSPSLIGSGIGFTIYOE-----IING-----IEPAKP-----AE 400
 Db 276 PGPKDYSLIGPSIALGNDGTGKWKQDDGYTSVNNGTFTFLSPSESTSLRKFVAGYST 335
 QY 401 AGTD-----ALPILAKDIEAEYKESGLNV-----GYSIYNTGEOSNLDGYRL 446
 Db 336 NGTDLTPTPTENALATVYTHYDNNNAFGDGLLGYOGGNVHHYRKGTTININHGSL 395
 QY 447 NKADGEVIEYKTSINI-----
 Db 396 LVTPG-IIDVIGSVNIDGRNNASTAFKNTGSSVDNMTISVWNTFTNPSEGRKN 454
 QY 464 -----SMYGEHPESEFSLAPNOLSGINTITLLYRBTGEOMEPEYRHAOG-----GY 511
 Db 455 VNEISDATSMSTIQR-----LTTGEVEMVNG--SESSGYTAGNRKVHTTGEISSA 506
 QY 512 VNSIKVNTDPNNVVTVDNNEKLSIVNPFADINSYEHSTITVQFNSDSPDEIRTPV 571
 Db 507 VNALRIWMDYG--VIFRSEGLHIITPAVGEKNG-----DIGPLR--PF 549
 QY 572 APLALSG-----ATADVISLGVN--MAEVPDSSNYPVYMSKVYVLTLSGD----- 616
 Db 550 STALDGVKVIIPDLESSYNTFAANGYIKFAGHAGAGGYDIOYQAAPIFOEIDDAVSK 609
 QY 617 YTLMTRESINNOCKEMKIGSVKTPREYRHPLEFVG--HNOTSTYTLDMANRYLDPDT 675
 Db 610 YPIYQKRLNGKAVN-SIGT-EINSQTYVHLHKEDSGSGHTSRPNADGTN--PPDNY 665
 QY 676 LKNLG--LPFNGELVVERQOGSSGSLMAADEVTHI-----KQGETFYVK 719
 Db 666 QVGGGEATIARNGNIRSDIMKFTTSAG-----ETINIDALATRVSKEDGDMTGK 715

Search completed: May 29, 2003, 13:26:02
 Job time : 22 secs

08rs9 escherichia
08rp2 methanococc
09796 thermoplasma
08rl58 methanococc
08rl75 methanococc
09y944 acroporum p
08py7 methanococc
09xb4 porphyromon
08ls7 methanococc
08z96 salmella
08xy4 escherichia
08rp2 methanococc
09l37 thermoplasma
024835 actinobact
09k197 vibrio chol
08rp1 methanococc
09h24 thermoplasma
08y9n listeria mo
09z66 listeria in
00686 escherichia
08n88 escherichia
08rl65 methanococc
09l907 caenorhabdi
0974r2 sulfolobus
094647 synecocystis
08rl44 methanococc
09rg3 escherichia
09wv5 gallus galli
087658 salmella

```

Db 74 YKAEEREE-----ALFEVFNNGKDGDFLLVAADRFPEVIGYAEFGHFAARMPD 123
QY 119 NLRMMLQIYDOEIGLILSGKAQLNEELR---TEGPAEYHALMDNGHFANDPMRNGY 175
Db 124 NLRGWLKGYREKLELWADGAERIDIPREKPTDRDLPSSAPLLELGEHNSDPLMDQY 183
QY 176 PMNNKEPRLPNHNAATGCAVATAAOIMRYHSPRLQEGSEFDYHAGSLVG---NMSGTEG 232
Db 184 PFMTLLPRLPSSGOAATGCAVATAMGOIMRHYKMPKASGEXDYD--DDMTGTHHYSGTFCG 242
QY 233 EMDWMINMPGNPLDNLTOGVDAYATLMDVDASASMSYENGSGTSYUYVGAALRNRF 292
Db 243 ETINMSKMPGNISV--GISPEEVALSTFMDVDSFVNMOPADSGSGTSLFEVERALRETF 301
QY 293 RYKRSLOLHYRALYTSOEHMDIRGELASGRPYVYAGNNOISGHAFVCDYASDGTFFHN 352
Db 302 HYKKSRLRYHRSLLPKCKEMDMIRKELAEHNPVYVYAGDSMGHAFVCDGTEPDGFHN 361
QY 353 WGGGYSNGPYKLTLLSPSLGIGEGIGFTTYOEITITGTEPAKTAENACTDALPLALK 412
Db 362 WGGGHSNGFYNLNLNPGSLGAGDGGYSTDEEVYIGIEPASNEA--PGIVDPPTITLY 420
QY 413 DIAEYKSEGLNNGSYVNTGEGSNLDGRLNKADEGEVIEVKTSSINISMYG--YGEH 471
Db 421 GLO-HMMSDALDLKSVKIKNYSTYAGDVKATLRLPNCGETTNPAVTVFVMEIDIGES 479
QY 472 PESFSLAPNQLSGINTILLYRTGTEDMEPVRAHOGYVNSIKYNTDPMNVVTVDN 531
Db 480 TGNITIPCSOFAEGKNTISILYRTDGMADKELKILMLGVNKIEV--TMPAGDVAVSY-- 536
QY 532 NECKLSIVPSFPAADLNSYHSTITVOFNSDSDPEIRTPAFALSGATADVDYISLGVW 591
Db 537 ADRIYALKDGLSHNKAAYSCKLSATVYVNPGEERSRPTFALRN--TEGRLYFGLRHL 594
QY 592 AEV-PGSSSYNYVWMSKDVLTLSGCDYTLWYR--FSINQKDEMKIKGVSVPTEYTH 648
Db 595 VELHPDDEDEKSLTITGKARAGCYMLVCTGDMELMADSMIELASIEVAHEHSTHS 654
QY 649 PLFEVGHNTSTYTLDMANRVLDPFTLKNL--GLPENGELVYVROTQSSGSLMAOET 707
Db 655 SLIVASNPQIDLLTVHANPELPTFSITNEGATFSKIEIYAKAFSET--FFQAKKEH 713
QY 708 VHIKQGETPYKRVVEGP-----IPDSYRATLHAFAVNGOOLY----LKGRNTY 754
Db 714 MSIAQGETKVLSEPLANSLSYTNALFPDGIYIYIR-----EOGFWDIDLEFGDYR 768
QY 755 VKIYNGTAAVAIESSSE--EIRVFPNPARDVEISAPCIQETSIILFDSGLKIVMKNSL- 811
Db 769 IRLITDSSDIDAGKDVSTVLYLPNPAHDYVHAIFPTVYAGSTLRLPFDIGRMQLSTIR 828
QY 812 -----SAGH 815
Db 829 ICRYASRRRTSSEGH 843

```

RESULT 2

```

ID 09AJD2 PRELIMINARY: PRT: 398 AA.
AC 09AJD2;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE Pyrogenic exotoxin B.
GN SPE B.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID-1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NZ131;
RA Watanabe Y., Ohkuni H.;
RT Cloning and expression analysis of Streptococcal pyrogenic exotoxin B

```

```

RT (SPE-B)/Streptococcal cysteine protease (SCP).";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB051298; BAB40954.1;
DR InterPro: IPR000200; Peptidase_C10.
DR Pfam: PF01640; Peptidase_C10; 1.
DR PRINTS: PR00797; STREPTOPAIN.
SQ SEQUENCE 398 AA; 43218 MW; 07D1F534E7867CDD CRC64;
Query Match 11.2%; Score 498; DB 2; Length 398;
Best local similarity 30.6%; Pred. No. 1,9e-21;
Matches 129; Conservative 74; Mismatches 154; Indels 64; Gaps 13;

```

```

QY 6 LLAIVMLFGITAMG-----HSAPYTERKALSLARLALROVSLRMGOTAVSDKISIDYV 59
Db 10 LLSILLGFEVLNLPYFADQNFARNKEKAODSAITFLQKSAIKAGARSAED--IKLDKYN 68
QY 60 ROGDAERGITSOEGSPAYFVYANRGNNGYALVADDDPRTPLAVSPIGRFPMDSMPN 119
Db 69 LGSEL-----SSNMVYINISTG---GVIVSGDKSPELLGISTGSGFDVNG--KEN 116
QY 120 LRMMLOIY-----DOEIGLILSGKAQLNEELITREGVPAEYHALMDNGHFANDPMRNG 173
Db 117 IASFMSYVEQIKENKKLDPTTYGTAELIQPV-----VKSLLD-----SKGIHYNQ 162
QY 174 GYPMNKEPILPN-----GNHAYTGCVATAAOIMRYHSPRLQEGSEFDY----- 218
Db 163 GNPYNLLTPYIEKKEKGEOSFYGOHAATGCAVATATQIKYHYPKKGKLDYTYLSSNN 222
QY 219 ----HAGSLVGNMSTGFGEYDWIN--MPGNPLDNLTOGVDAYATLMDVDASVSMSPY 273
Db 223 PYRNHPRNL---PALSTQIYNNNLLPYSGHSNVOKK--ALSELMDVGIQVMDYG 277
QY 274 ENGSGTYSYVYVGAALNNFRKRSLOLHYRALYTSOEHMDIRGELASGRPYVYAGNNS 333
Db 278 PSSGAGSSRYQALKENKCYNGSVHQINRSDPSQDKMDSQDKELSQMPYVYQGVKY 337
QY 334 IGHAFCVDGASDGFHFPMKNGCVNGYKLTLLSPSLGIGEGIGFTTYOEITITGIE 393
Db 338 GGHAFYIDGADGRNFHYWVGWSDGFEFLDALNPALGIGGAGGNGVQSAVYGIK 397
QY 394 P 394
Db 398 P 398

```

RESULT 3

```

ID 093IQ2 PRELIMINARY: PRT: 344 AA.
AC 093IQ2;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Cysteine protease SpeB (Fragment).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID-1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SP268;
RA Jadoun J., Eyal O., Sela S.;
RT "Role of cstr, hasa, and speB genes in the internalization of Streptococcus pyogenes by epithelial cells."
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY035886; AAK71464.1;
DR MEROPS: C10.001; -.
DR InterPro: IPR000200; Peptidase_C10.
DR Pfam: PF01640; Peptidase_C10; 1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 344 344
SQ SEQUENCE 344 AA; 37345 MW; FB0A9BD05B541909 CRC64;

```

Query Match	10.48;	Score 461;	DB 2;	Length 344;
Best Local Similarity	31.38;	Pred. No. 2.3e-19;		
Matches 116;	Conservative 66;	Mismatches 131;	Indels 58;	Gaps 12;

```

0Y 38 ROVSJRMQOTAVSDKSIDIDYVYRQGBAERGITSQEEGSPAYFVANGNNEGVALVAAD 97
0Y 4 KSAATKAQARMSAD- IKIDKVALGEL-----SGSMNYIYNISTG--GVIYISGDK 51
Db 98 RIPTILAYSPIGRFDMDSMPDLRLMWLOIY-----DOEIGLILSKRAOLNEBILRTGEV 151H
Db 52 RSPELLGXTSGGSPDANG- KENIASMESYEQIKENKKLDITTYAGTAIEIKQV----- 104H
0Y 152 PAEYVALDNGHGFADPMMRQNGYPRNNKEPLPN-----GHNATGCVATAAQI 202H
Db 105 ---VKSLLD-----SKGIHYNOGNPNYLLTPIEYKVPGEQSFVGQHAATGCVATATAOI 156H
0Y 203 MRHSMPLOGEGSFY-----HAGSLNMGSGTFGEWMDWIN-MGPNDDLIDLIQ 251H
157 MKYHYHPKNGKLDIYTTLLSSNNPYRHPKNL---FAIISTROYNMNNILPTIYSGRESNVQ 213H
252 SQVDAYATLMDRDVASVSNMSEFYENGSGIYSVYVVGALRLNNFRYKRSLOLHVALRYLSQEW 311H
214 KM- AISELMADVGISVMDMDYGPSSSAGSSRVRQALKEHFCVGNQSVHOINRSDFSKQD 271H
0Y 312 HDMIINGELASGRPYTYAGNNOSIGAHFVCDGYASDGTFFHPNMCGMGVSGNGFYTLTLLSPT 371H
Db 272 EAQIDKELSONOPVYVYOGVKGVGHAFVIDGADGRNFRFHYVMWGMGVSDGFFRIDLALNPS 331H
0Y 372 SLGIGEGIGF 382
Db 332 ALGTGGGAGGF 342

```

RESULT	4
P72198	
ID	P72198
AC	P72198
DT	01-FEB-1997 (TReMBLrel_02, Created)
DT	01-FEB-1997 (TReMBLrel_02, last sequence update)
DT	01-DEC-2001 (TReMBLrel_19, last annotation update)
DE	Hemlin-regulated protein.
GN	HEMR.
OS	Porphyromonas gingivalis (Bacteroides gingivalis).
OC	Bacteria: CBH group; Bacteroidetes; Bacteroidales; Porphyromonadaceae.
CC	Porphyromonas
OX	NCBI_TaxID=837;
RN	[1]

[illegible]

```

QY 559 FNDSPDELRTPAFAFLSTGATADDDYISLGMWMAVPBGSSNIPVWSKVDLTSEDYT 618
Db 194 -----ELHPDEDEGEVSL--TITGLKRAAQOIMVCTGDMESLME----- 222
QY 619 LWRFSINNOKDEMKIGSVKPTLEYTHPLFEYGHNOTSTYTTLDMANHNVLDPFTLKN 678
Db 233 -----DASWIELASIEVAEHTSTHSSLLVNASNQIDLLTVHRANPETTLPTFSITN 288
QY 679 L-GLPENGELVYVFROTOSSGSLMAAOETHYIKGEPFFYKRPVECP-----IPD 728
Db 283 EGATSSGKREIYAIATASET--FPOAKERHMSLMOGETKVLSPELTANSSLYTMAELPPD 344
QY 729 GSYRATLHAFAVMSQOOLY---LKGKRNTRYIVNGTAVEAIESSE--EIVRFPNPARDY 782
Db 342 GIYYIYIIR-----EOGFMPIDILFGDYYYRIRLITLDESSDIAGKDVSTIYLPNPADHY 396
QY 783 VEISAPCIPOETSIILFDLSGKIYIM 807
Db 397 VHYAIPPTYAGSTLRLEFDIGRMYI 421

```

```

RESULT 5
09S3R0
ID 09S3R0 PRELIMINARY; PRT; 540 AA.
AC 09S3R0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
DE Immunoreactive 61kD antigen PC91.
OS Porphyromonas gingivalis (Bacteroides gingivalis) .
OC Bacteria; CFB group; Bacteroidetes; Bacteroidales; Porphyromonadaceae;
OC Porphyromonas .
OC NCBI_TaxID=837;
OX [1]
RN
RP
RP SEQUENCE FROM N.A.
RC STRAIN=MS0;
RA Ross B.C., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "Porphyromonas gingivalis polypeptides and nucleic acids." ;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBU databases.
DR EMBL, AF175723; MADS1076.1;
SQ SEQUENCE 540 AA; 61128 MW; 7F4E9446C0C38966 CRC64;

```

```

Query Match Similarity      4.3%: Score 190; DB 2; Length 540;
Best Local Similarity      20.5%: Pred No. 0.0038;
Matches 103; Conservative 78; Mismatches 162; Indels 160; Gaps 23

QY      DIEAEKSESGINV-----GY-----SIYNTGEESNDLGRLNKADGEVLEVT 458
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      LDIYVTKNESDIKIMSVELMKKPGGYKSKVANVFSDANQNMKIYKEDFSNVQLYVDI 160

QY      SSINISWYGYEHPESEFSLPADNLSOGINTITLLYRTGTEOMEPEVYRHAOGYVINSKVN 518
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      161 AS-----NY-----RSPSLNNGGNPFLAEAYTGF----- 186

QY      TTDPNNVVYVDNNEGKSLIYPNISVYADLNSTEHSTIYQFNSDSPDEIRTVAFALSTG 578
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      187 -----NNHKKISFV---DVFSLNGGONKNKLLFSODEKKID---KYDLSLG 229

QY      ATADDVISLGMVAEY-----PGSSNVPV-----WS-----K 607
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      230 STSESMGINNAPFLMGVFEEMKQGGKSDLGFLSNFYVNDPEFQWSPRIYVESDMSFSK 289

QY      DVLTLSEGDYTL---WYRFSL---NNQKDEMKKIGSVSKTPLEYTHPLFEVGHNOT 658
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      290 IQMLLDEEDNNINSGESCHNEMITTYSDIYSEYSDW-----DIRYVPPKSFYEKG 339

QY      STYTLDMANRYLPDF--TLKNLGLPF---NGELVVVVFROTQSSGSL---MAAOETVH 709
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      340 KTRPMDLDVEAFILFASYSQETNSGIGYKRNANHUYLITAAKKKEENGTNLTLYRWANYDKIH 399

QY      IKQ---GEYF-----YKRPVEEPIPDGSTRALILHAFVNCQOL-----YLGKRN 752
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

```

Db 400 NKDLMSPTFTYTSSANLTPPOVD-----INPTKGLVCMSWEYLPGKRI 444
 QY 753 Y--TVKIVNGTAVEAIESSE-EIRFPNPARDVEISAPCIPOETSIILFDLSGIYMKN 809
 Db 445 VMSDQTHANGVEDIYMOGSMKLYPNPAQEVAVISLP--TAANCKAVYVDMQGRVAAE 503
 QY 810 SLASGHRMRYSRPLNCAIYILKV 832
 Db 504 SFGSNEYRLNVOHLAKGTIYILKV 526

RESULT 6
 ID 082411 PRELIMINARY; PRT; 3624 AA.
 AC 082411;
 DT 01-MAR-2002 (TREMblrel. 20, Created)
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 Large repetitive protein.
 STY2875.
 Salmonella typhl.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC NCBI_TaxID=601;
 OX NCBI_TaxID=601;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaiha M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltywell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrell B.G.;
 RA "complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhl CT18.";
 RL Nature 413:848-852(2001).
 DR EMBL: AL627276; CAD05867.1;
 DR InterPro: IPR003410; Hyalin.
 KM Complete proteome.
 SQ SEQUENCE 3624 AA; 365353 MW; BB52E561D6998F14 CRC64;

Query Match 3.9%; Score 173; DB 16; Length 3624;
 Best Local Similarity 20.3%; Pred. No. 0.64;
 Matches 223; Conservative 102; Mismatches 376; Indels 400; Gaps 54;

14 G1AMOGSHAPVTKERALSLARLALROYSLMGOTAVSDKISIDYVROGAERGIR----- 69
 2545 G1A-AGFTAPDTRVPEAPITTVDDVGYTGAINGQVTDNAQPTLNGTAQAGATVSIY 2603

QY 70 -----SOEGSPAYFYVANRGNEGYALVAADRIPTILAYS 106
 Db 2604 NNCALLGTTTANASGNMSTFTPTGNLTGESHAFATATNANGTSVSTAATVIYDILAPGT 2663
 QY 107 PIRCFMDSKPNLRLMLQIYDOEIGLISGAQLNEEIIRT--EGVPAEVHALMDNGHF 164
 Db 2664 PSCGTLISAD-----GGSLSGQAENSTVYVTLAGVTLITTTA--GSGN-- 2703
 QY 165 ANDPMRNGGYPMNKE-----PLLP---NGNHATGCVARA 198
 Db 2704 -----AMSLTLPKQIIEGOLINTVATDAAGMASGTLGTAPILPLAARDNITSLLDTSRA 2758
 QY 199 AAQIMRYHSMPLQEGSGSFYHAGSLVGN-----WSGTGEGM----- 234
 Db 2759 VNSTQVSYDYGILLVVALG--NVASVLCNGDTAQVEFTIAGGTCGVYIIDAATGIVLSLS 2817
 QY 235 -----YD-----WINMPGNPDLDNLTSQSDAYATATLMRDVSASVMSFENGSGTYS 281
 Db 2818 TOEIVQVRDTSIGTY-----TTIYVTAVGDEANLLTLTLGSGVTILNLNGLGGQYR 2868
 QY 282 VVVVGALRNFRKRKSLQHLVRALYTS--QEWHDIMRGELASGRPVYVAGN----- 330

Db 2869 V-----LYNTSILL--ATGSYTSILDVYHQTSAG--IISG--PTISGNVYADDTAPT 2915
 QY 331 -----NOSIGHAFVCDGYASDSTFHPNMWGVSGNCFKTLILSTSGIG- 376
 Db 2916 GTTVVATATNANGVSTPVGAGV--DILGQVTLIN-----QGSYTYTLTKPTA--GYGH 2967
 QY 377 GEGIGFTIYQ-----ELTNGIEPAKTPAEA--GTDALPILAKDIEAEY-----KSES 422
 Db 2968 KESFTYTIINGVGSSAAQLYINLGPAPVPGSVIATDNNASLVF--DTHSVYVNGSPSTGS 3026
 QY 423 G---LVNGVS-----LYNGEEO-----SNLDL--GYRLN 447
 Db 3027 GVTVLSVGLGNVLNANLMDMTNPILIFNVEEGATRTMTLQGTGVGSIVSTEDLYVYRPN 3086
 QY 448 KA--DDEVIEVKTSINISWYGEHPEBSLAPNOLSGINNTITLIRRTGEQPEPR 505
 Db 3087 DAIOQYEQFRVQGWGINTLILLAGOSQPLTTLTGEGEYLFVNLASGISVLTGYTLAISOD 3146
 QY 506 HAQGYVNSIKVNT-----TDPNNVNV----- 527
 Db 3147 HTVA--VDSITANTGNVLTNDVAPFDALLTEVNGVAIAATGTEVNGLSLIDARGN 3204
 QY 528 ---TVDNNEGKLSI--VPNSFVADLNSYESTITVOFNSDSPDEIRTPVAPALSTGATAD 583
 Db 3205 YTYTLKNGVAGDSIKTPDSFIYTLKAPNGGDTVASINI-----TPFARALDAINDVSD 3257
 QY 584 VISL-----GWMAEV-----PGGSNNPVYWSK-----DVLITSEBD 616
 Db 3258 TISVATLQDTAMALDSSVGSAGMLKSGSGSGTDFVATGYLAKASLVFDVSTL----- 3313
 QY 617 YTLMYRFSINNOKEWK-----KIGSVYVTPREYTHPLFEVGHNOTSYTLDMANH 668
 Db 3314 -----ITLGNLMSAIDENGTVINGVYPVANIT-----LG--SAYTVNLS-- 3354
 QY 669 RVLPTDTLNLGLPEF--GELYVVFROTQSSSGSLMAOE-----YHAKGCFPVYKPYVE 723
 Db 3355 -----GLELDAGTYTLNFTGTNLTAGATTPHYIGTYVLDLDFEFTSGTHTVL- 3402
 QY 724 GPIDGS-----YRATLHAFVN-----GOOQLYKGR 750
 Db 3403 GNIFDSDDAAGAMDQNTVYNTRLISIGYNSAATLDAANTISATIQGHYGLQIHLDA 3462
 QY 751 RNTYKIVNGTAVEAIESSEIRFPNPARDVEISAPCIPOETSIILFDLSGIYMKNS 810
 Db 3463 YFTYTLN--NGVAMSSITSKE--VFYQLDDKRIGHT-----DSATLTIDMAPQIYSTNQ 3511
 QY 811 ----LSAGHRMRYSRPLNCA 827
 Db 3512 NDVLISAVGDITLHYHLNGA 3532

RESULT 7
 ID 093417 PRELIMINARY; PRT; 1330 AA.
 AC 093417;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE MS115, putative beta-agarase.
 OS Micrococcilla sp. PREL.
 OG Plasmid PSD15.
 OC Bacteria; CFB group; Flexibacter group; Micrococcilla.
 OX NCBI_TaxID=155537;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=PREL;
 RA Zhong Z., Toudkardian A., Helinski D., Knaut V., Sykes S.,
 RA Wilkinson J.E., O'Byrne C., Shea T., De Loughery C., Caspi R.;
 RT "Sequence Analysis of a 101 kb Plasmid Required for Agar-degradation
 of a Micrococcilla strain";
 RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF339846; AAK62837.1; -.

DR InterPro: IPR003343; Big_2.
 DR InterPro: IPR000847; HTH_LySR.
 DR Pfam: PF02368; Big_2; 3.
 DR PROSITE: PS00044; HTH_LySR_FAMILY; UNKNOWN_1.
 KW Plasmid.
 SQ SEQUENCE 1330 AA; 143942 MW; 083073D8BFCB28F CRC64;

Query Match 3.7%; Score 166; DB 2; Length 1330;
 Best Local Similarity 20.0%; Pred. No. 0.37;
 Matches 110; Conservative 101; Mismatches 337; Indels 244; Gaps 37;

QY 91 ALVAADDRPTLLAYSPGRFDMSPDNLRLMLQIYDDEIGLLSGAQLIEILLTEG 150
 DB 614 ALYNDOD-----NOVGHLSIDKLTIVNT-----EITPLDMGTNGVRRNYFSTLE 659
 QY 151 VPEVNAADNGHAFADPMRNNQGYNNKKEPLLPGNNAHYGCVAATAAQMIRHSWPL 210
 DB 660 IIPVGLLTQNNFTYD-----FHHVEVAVVNIQTWEFSKVP 658
 QY 211 QGEGSPDYHAGSLVGNMGTGFGAMYDWINMGNPDNLNQSQVDAVATLMDVSVSM 270
 DB 699 RSTSSDPVAVTGYVS-----PGSPYAQGSTVQWTAHVQPTNATDOSVTW 744
 QY 271 SFYE-----NGSGTYSVYVVG-----ALRNFRYKRSLOLHY-----R 303
 DB 745 SSNNANVAVSNAGCEVTGIAQGTATITATNDGFIASQVWVTTGDVDTGVSVPTSA 804
 QY 304 ALTTSC--EWMHMRGELASGRPYTAGNNQST-----GHAFCVDCVATSDGTFHNMWCG 357
 DB 805 SILVGSIDLTETVSPNTADKSVTWTSSAVYVNGSLVYAKNGSAYVTYTTNDG 864
 QY 358 VSNQFKLLTLPSTSLGIGEGIGFTYQELITGI-----EPAKTPA-----EAGTDALP 407
 DB 865 -----FTQSNVYTTTGGGSAIYEAESFTSTGTTDSDPYGGGIGIVNNAGIN--- 914
 QY 408 ILALKITEAYKESGLNVG---YSI-YNTGEOQNLGLYRLNKAAGEVIEYKTSIN 462
 DB 915 ---INVNSBMAEYGINVSEAGTYQIEYQISTPNSNAQVREFEL---DGNVSTDNVFN 968
 QY 463 ISMYGGEHESFSLAPNLSQINTTLLYRTGTQMEPVRAHAGGYNSIKVNTDP 522
 DB 969 GQMDSTKLIASTI--STLSTGSHYRLVAGANAMQW-----NLKRYLTJKT 1015
 QY 523 NNWVTVVNDNEGKLSIYVNSFVADLNSYESTITVQFNSDSDPEIRTPVAFALSTGATAD 582
 DB 1016 GS-----STVAVTGVSAFPTNV-----SLISGSTD 1041
 QY 583 DVISLGMVAEVPDGGSSNPVWMSKD-----VLTISEGDTL 619
 DB 1042 LLETIVN-----PGNATDKSVMSNNTAVATVDANGLVAVGAGTAVITVTSDDGHTA 1095
 QY 620 WYFESINOKDEMKKIGSVKPTETHTHLEFY-----GHNOTSTYLLDMAHNR 670
 DB 1096 TCSVTVVSGNSVELTQADEPAT--TGCTHDGFQYVAVGATATNMGTGMA---DYSVT 1151
 QY 671 LP---PFTLKN-LGLFENGELVYVFRQTOSSGSLMAAOETVHI--KOGTFYVKKPVYEGP 725
 DB 1152 IIPAGDYSTIEYFAGTYVNGAAVTI-----SYDG---AVQRTDAVPNNGNMDFESLAVGG 1203
 QY 726 IPDGSYRATLHAFAVNGOQOYLKGRNY-----TVKIYNGTAVEAIESS-----EE 771
 DB 1204 -----RISLNA-----GSHTRILGDDGTNGEMNMNDREVLSKGAASSVESSSQIYNQG 1254
 QY 772 IYVFPNARDYVETIS--APCIFQETSILFDLSGKIYKNSLSAGHR--MDVSRPLNGAY 828
 DB 1255 ISITVPPADKRIYVRLAPDLQLT---ISVSGKIYRKMSVEGPNDIIDVGLKTYGV 1311
 QY 829 ILKYDGYTKIN 840
 DB 1312 FLHFHSGKTYFN 1323

RESULT 8

093416
 ID 093416 PRELIMINARY; PRT; 614 AA.
 AC 093416
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE MS116, putative beta-galactosidase.
 OS Microscilla sp. PREL.
 OC Plasmid PSD15.
 OC Bacteria; CFB group; Flexibacter group; Microscilla.
 NCBI_TaxID=155537;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PREL:
 RA Zhong Z., Toukadarian A., Hellinski D., Knauf V., Sykes S.,
 RA Wilkinson J.E., O'Brien C., Shea T., De Loughery C., Caspi R.,
 RT "Sequence Analysis of a 101 kb Plasmid Required for Agar-degradation
 RT of a Microscilla strain."
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF339846; AAK62838.1; -
 KW Plasmid.
 SQ SEQUENCE 614 AA; 67878 MW; A9571EC7D32A7793 CRC64;

Query Match 3.7%; Score 164.5; DB 2; Length 614;
 Best Local Similarity 19.4%; Pred. No. 0.14;
 Matches 138; Conservative 82; Mismatches 262; Indels 229; Gaps 32;

QY 222 SLVGNMGTG--EYMDIMNPNPDNLNQSQVDAVATLMDVSVSMSTFENGSG--T 279
 DB 22 TLITWTAFAAGAAQODGIP-----VPASA-----GGGKT 52
 QY 280 YSYVYVGAALNNFRYKRSLOLHYALVYTSQEMHDMIRGELASGRPYVAGNOSIGAFV 339
 DB 53 WQLQASADPFTNFETNSQL---TNFSKKNWTA-----FY 85
 QY 340 CDGYASDGTPEHFMWGWGVSNGEYKLTLL---SPTSLGIGEGIGFTYQELITGI 396
 DB 86 HNGMDGDTGYWQYNNHVSAGNGLVLRASNPSTTKMGVPGVNAV-----CITSNNRVK 139
 QY 397 TP--AEAGTALPLALKDL-----EAETSEGLNVGY-----STY 431
 DB 140 YPVFVEASVAVANALASDVWLLSPDPTQIDIIIECYGAGSNNAVPAQFIHLSHSFVR 199
 QY 432 NTGEOQNLGLYRLNKA-----DOEVEYK 457
 DB 200 NPFODIOPRLNLSMKKSGVSMGDIYCMNNKRYVAVGNWPKFEYIIDGELVRL 259
 QY 458 -----TSSINISWY-----GYGEHESFSLAPNLSQINTTLLYRTG 497
 DB 260 YDKAFATKAVGTWYTYPTMTNGSLDPSGGY---QSVQYATGSSYSFSLQASNTSS 315
 QY 498 TEOMEVYRAHAGGYNSIKVNTDPNNVYVVDNNEKLSIVNSFVADLNSYESTITV 557
 DB 316 TSVYIDPYNV--QGGN-----GFTKELDIIINVEQNMHVRGRTPSDADINDPAKKNMKV 368
 QY 558 QPNSDSDPELRT--PVAFALSTGAT-----ADVVISLGMVAEVPDGGSSNPVWMSKV 609
 DB 369 -----DWLRVTKPVYTTGGGGSASITYEAESFATGTGYNDASSGGGEGYKAKGTI 421
 QY 610 LITISEGDTLLMYFESINOKDEMKKIGSVKPTETHTHLEFYGHNOTSTYLLDMAHNR 669
 DB 422 INYVNG--CQWVEPFI-----VPEE---GVYEGVH---YATPISNTS 457
 QY 670 VLDPFTLKNIGLFPNGELVYVFRQTOSSGSLMAAOETVHIKOGTFYVKKPVYEGP 729
 DB 458 V--QFSVS--GTFP-----FTTTLTNTGG--MGNTQTRASQTAFT-----AG 495
 QY 730 SYRATLHAFAVNGOQ-----QYLKGRKNTVRYVNGTAVEAIESSSEIRVFPNARDY 783
 DB 496 NHLKLTLAGSADQWQNMNDKVTLTRVGNMNSRYLIVPTSDVEGY---DIVVFPPTTDDL 551
 QY 784 EISAPCIQFETSILFDLSGKIYKNSLSAGHRMDVSRPLNGAYILKVDG 834

Db 552 RI0G-LDDKMYVYLYDLGNTNYSROYKGAERLVDYSDLSGILYLLKLEG 601

RESULT 9

08KX8 PRELIMINARY; PRT; 1588 AA.

ID 08KX8

AC 08KX8

DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Stalidase-like protein.

GN CpeI254.

OS Clostridium perfringens.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;

OC Clostridiales; Clostridiaceae; Clostridium.

OX NCBI_TaxID:1502;

FN 11

SEQUENCE FROM N.A.

STRAIN-13 / TYPE A;

PUBMED-11792842;

Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A., Shiba T., Ogasawara N., Hattori M., Kubara S., Hayaishi H.: "Complete genome sequence of Clostridium perfringens, an anaerobic flesh-eater";

Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).

EMBL; AF003189; BAB80970.1;

DR InterPro: IPR000421; FA58_C.

DR InterPro: IPR003961; FM_III.

DR Pfam: PF00041; fn3_1.

DR SMART: SM00060; FN3_1.

KW Complete proteome.

SQ SEQUENCE 1588 AA; 178194 MW; D34C534B41E04D49 CRC64;

Query Match 3.7%; Score 164; DB 16; Length 1588;

Best Local Similarity 18.0%; Pred. No. 0.63;

Matches 206; Conservative 164; Mismatches 361; Indels 412; Gaps 58;

QY 2 KKSFLAIYVLFGLMAGHSAPYTKERALSRLARLRQVSLRNGQAVSDKISI--DYVY 59

DB 8 KKLISLTL-----GLSVALTNTPAPLSVLANGAKVEDVAVSKS--QDKITIGNEYIS 56

QY 60 RQGDABERG--ITSQEGSPAYFYVANRGNNEGVAL--VAADDRIPIILLASPIGRDMS 115

DB 57 RESEIENGKVLTEISEIENSRANTTLVLPOLGSEDTITITQENSDLPVEANSP----- 108

QY 116 MPDNLRMWLOIYDOETIGLLSGRAQLNEELIRTEGV--PA-EVHALMDNGHFANDPMRN 172

DB 109 -----KEVLD-----RANNNALITLANSCTAYATDIEKLLDDGKNYIDSYNI 151

QY 173 QGYPMN-----NKEPLLPRGNHAYTGCYATTAQAQIMRY-----HWPILQ 211

DB 152 TGIPTSLKIDLGKVTSSFSYQKRQFTDANGKNGTM---GQYKLYSEDEGENWTEA 207

QY 212 GEGSPD-----YHAGSLV--GNNSGFEGEMV----- 235

DB 208 GEGEFREDENLHQEENLHNVGDVYGNDKYTESKXIRIDQLSDALGTOERSCAEINL 267

QY 236 ---DWIMPGNP-----DLNLTQSOVDAYATLMDV---SASVSMSEYENGSGTYSV 282

DB 268 YSDKYKEPEEPIAPKAIKSSDLT---IDTSFKIEDIENGKLLTISYPEYENGTEYNI 324

QY 283 YVVGALRNFRYKRSIQQLHVR-----ALYTSQEHMDMT--- 315

DB 325 DMVTVLENDQHYRNF--LEIKTNENENAKIDYIDLHFILEDGIKIDTLMSHPLEEDVTSIL 383

QY 316 --RGLASGRPYVYAGNNSIGHAF-----VCDS-----YASNG 347

DB 384 IKNELMLGQPIY--ANGMFSEFPATDTDVDDQMORRYSGKSFEELENDQLTDDG 441

QY 348 THFNM--GNGVSNCF-----YKLTLLSPTS-----LGIGGEGIGFT 383

DB 442 KF-VSMQNVYGAAGVDTDVQTDFFEFYISEIATPTDFRKQYNSWDMNLITIDSISIS 500

QY 384 IY--QELIT--GLEPAK-----TAEAG-----TDALP- 407

DB 501 FYGTENGLENGVEPEIDSYVDDGMHNYRDPPEPNPISKEQAGNSMRTGFMEFNKFPN 560

QY 408 -IILAKDIEAEYKESGL-----NVGSIY-----NTGEQSNLDGRLNKADE 452

DB 561 ELTSTELNISKPSKSGMLMGPQSGINFGFARYLEKMGTYAQTN--NGNVVCGSR 618

QY 453 VIEVKS-----SINISWYGGHEPESFSLAP-----NLSQGINTI---TLLYRT 496

DB 619 YIKNLTSFLDYOKRFDIDYWKL---DGFALPCTSENHDMHTGHNMYTTDLMEK- 673

QY 497 GTEQWEPVRAAGGYNSIKVNTDPNV-----VYVDNNGKLSIYPSNVADL 547

DB 674 WTMAMETMARSAEKGOLFYNATCYVNSPWLQVNTVWQNSO-----DTGEAGT 726

QY 548 NSEHSTIT---VOFNSDSPDEIRPV-----AFALSTGATADVISLGWVAEVP 595

DB 727 GSRHQOKITRYDAVYHDIYKSNQIOFPKNTYNHPEIYGVSDGFATFEDFPLFANV 786

QY 596 GGSNNPVVMS-----KDVLTLSGDDYTLMTYFSTNNOK-----DEWK-K 634

DB 787 RGTAFELWLYSPSINDERKKNVADLVPEVNNFNLEKAKLFGHRTGEGYYSAMGN 846

QY 635 IGSVYKPTPEYTHPLFEVGHNOTSYTIDMAH-----NRVLPDETLKNGSL 681

DB 847 EGIVSFRNPT-----GETKRTIDLDIDYGVPSVNLGNQVLP-YKVGDDGS 894

QY 682 PNEGELVVVFPROTQSSGSLMAAQETVHIKQGETVYKPVGPIPDGSRATLAFVNG 741

DB 895 VSYGDSITVTLF-----PYETRILOYGKV-----DNKACEIVSAKVTG 932

QY 742 QOOLYKGRKNTYKIVNGTAVEAIESP-----EIVFPNP 778

DB 933 ENBITIK-----INERENEEYVSEGNITFTKLLDRTVYISTSKNKGAEALNNG 988

QY 779 ARDYVEISAPCIPOETSI-ILFDLSGKIYMKNSLSAGHGRMDVSRPLNGAYILKVDGYT 837

DB 989 EBDALQ-----NPLTSLTIPYNNKGIISVSGSELVGENINKKYN-----NSDIYFL 1039

QY 838 KIN 840

DB 1040 ELN 1042

RESULT 10

058048 PRELIMINARY; PRT; 1155 AA.

ID 058048

AC 058048

DT 01-AUG-1998 (TREMBLrel. 07, Created)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE 1155AA long hypothetical PYROLISIN.

GN PH0310.

OS Pyrococcus horikoshii.

OC Archaea; Euryarchaeota; Thermococci; Thermococcaceae;

OC Pyrococcus.

OX NCBI_TaxID:53953;

FN 11

SEQUENCE FROM N.A.

STRAIN-OT3.

MEDLINE-98344137; PubMed-9679194;

RA Kawarabayashi Y., Sawada M., Horikawa H., Hatake Y., Hino Y., Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y., Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamaya M., Ohnuku Y., Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Ioshizawa T., Nakamura Y., Robb F.T., Horikoshi K., Masuchi Y., Shizuya H., Kikuchi H.;

RT "Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";

EMBL; AP000001; BAA29383.1.

DR InterPro: IPR000668; Peptidase_C1.


```

QY 265 --SASVMSFEYNGSGTYSVYVVGALNNRFRKSLQHLRAL----- 305
DB 522 APDEEVSMDYTTHTGTAGIDYTPVYVNGSVYVSTGIDYRALNRTTGALNMTETIPDCS 561
QY 306 -----YTSOE-WMDMIRGELASGRPVYVYAGNNOSI-- 334
DB 582 GFOLGNPAYGNKIFVPTADSRIFAFDAMGTGELMDSVESNPKSSSLVHOJNTPVYVD 641
QY 335 -----GHAFCDDGASDCTFHFN-----KMGGSYS----- 359
DB 642 NGRIFYGOWIPANAGKYC--YTEDGETVWNRDCTNGNNGYVAGATVVDLYVEGGDD 699
QY 360 ---NGFYKL--TLSPSTSLGEGIGFTIYO-----ITGIEPAKT-- 397
DB 700 GLTSLYKVNGLTDEINIG-AGEIRSTLYDESNNRIFSTKNNGYISLEMMDEGFND 758
QY 398 -----PALAGDALPI-----LAKDIEAEYK--SEGLNAGYIYNT-GEOS 438
DB 759 SSKVQASAFDASSTFPIVINGRYVIGSGNFOASSPFCNSTDLSTVMTFTGTNGIGS 818
QY 439 NLDGVRNLKADGEVIEVKTSSINIS-----WYGGHPSFSLAPN 480
DB 819 SPVYSTAYDDGDEVTYFTTNNARYSLYCLKDNATNTNMTSEBT--FQPED--SKN 872
QY 481 QLS-QGI-----NTTLLYRTGTQEWEPVRAAGYVNSIKVNTDNNVYTV 529
DB 873 QYTLQVAVLVGMVLYGNDKYLEGLAGVONGSEIPAD--FNAAPLSGNSPLTVFT- 928
QY 530 DNNEGKSIYPNSEVADLNSEHT-----ITVQNSDSPDEIRT-- 569
DB 929 DSDTGS--PTSMFMDGDTNSTEONPSYTYTSAGNTVNTLVANAAASDSEVKDYI 984
QY 570 ---PVAFALSTGATADDAVSLGVMVAEVPGGSSNPVYVMSKD-----VL 610
DB 985 SVSEPLPSAPVYANFADITSGVGPLAVQFTESSTGFTANEMDNDGANDSTEQNFQI 1044
QY 611 TLSEGDYTLWYRPSINNOKDEMKKIGSVYKTP-----TEYTHPL----- 650
DB 1045 YTEGEGYSVKLTVYNSGDNSELKIDYISVYRPSDFEIANVTEGTVLVEFTGNVSGLA 1104
QY 651 -----FEVGHNOT--STYTLDMAHN-----RYLPDEF 675
DB 1105 PDVWMMAMVDVDDGTYDVSPNEFTYTYTEPGTYDVAVVADGTPNVKAGYITVLEIP 1164
QY 676 LKMLGLPEFG--ELVYVFRQTOSSGSL--WAAOE-TV-HIKOGEFYVYP-- 720
DB 1165 VANSADVYTGLEPLTVNF--TQSGEIVSYSMDFPNDGTVDISIBNPSYTESAGNT 1222
QY 721 -----VDEGPIPD--GSTRATLHA-----FVNGOOLY-- 746
DB 1223 VSLTVSNAGSDSELKTDYIVVSEPYEAPVANTATPTAGNAPLVTFNFDQSEGIYSY 1282
QY 747 -----LKGKRYTVKIYNGTANVLALESESEBEIVFPNPARDYEIS 786
DB 1283 SMDPNDGANDSTEQNFSTYESAGNTVSL--TVSNAGGSSELKT-----DYTVS 1333
QY 787 A-PCIOPESTIIILFDSGKIV-----MKNSLSA-----GHRMDVSRPLNGAVITLK 831
DB 1334 STVEPEPVAFTADYTGTAFLVNTDOSTGSPTAMANDFGDGANSTEQNFSHYV-S 1392
QY 832 VDGYYTKINT 841
DB 1393 AGNYTVNLTV 1402

```

```

RESULT 12
Q9X6M3 PRELIMINARY; PRT; 1605 AA.
AC Q9X6M3:
DT 01-NOV-1999 (TEMBUREL 12, Created)
DT 01-NOV-1999 (TEMBUREL 12, Last sequence update)
DT 01-DEC-2001 (TEMBUREL 19, Last annotation update)

```

```

DE Proline/threonine-rich protein.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_Taxid=601;
RN [1]
RP SEQUENCE FROM N.A.
RA Stojiljkovic I., Heffron F.;
RL "Salmonella typhi large orf 3' of smbP.";
DR EMBL: AF139631; AD34846.1;
SQ SEQUENCE 1605 AA; 162159 MW; 6230BF0803348377 CRC64;

Query Match 3.6%; Score 160.5; DB 2; Length 1605;
Best Local Similarity 20.3%; Pred. No.: 1;
Matches 183; Conservative 84; Mismatches 271; Indels 363; Gaps 47;

QY 45 GQAVSADKISIDYVROGDAERGIT-----SOEGSPAYFYVYVANGNNEGTA 91
DB 763 GLIVARDGLSV-----SGQAEAGATVYTRDSNTVLGSVAVANGNGQFIYPLNAAQTNGQA 817
QY 92 LVAAADRIPITLILASPIGRFDMDSMDNL--RMTQIYQOETGLISGKAOLNEELRT- 148
DB 818 LIAT-----ATDANNSSAAATYDA--PDSTAPEKKNVYISEDASISGTAPEGSSITITT 872
QY 149 -EGVPAEVHAMDNGHFANDPMKNOGYPMNKEPLP--NGNHAVTGCVATAAQMIR 204
DB 873 PDGTPLGSGKADGEHFT-----LPLAPQNTGEQ-----VYVATDSAN 912
QY 205 YHSPVLOGSGSPDYHAAGSLVGNKSGTFGEKYMIMNPGEDDNLNLTQSOVDAYATLMRDV 264
DB 913 NWSPTTAGP-DITR-----PDKPIITQ-----VLDPV 940
QY 265 SASVMSFEYNGSGTYSVYVVGALNNRFRKSLQHLRALYTSOEHMIRGELASGR 324
DB 941 ES-----FTGPLYNG-----QTDNRRPTLSGTAEAGAR 969
QY 325 VYVYAGNOSIGAFVCDGYASDGTFFHNMGW-----GVYNGFYKLTLSPTSLGIG-- 377
DB 970 VEIFDNGVSLGLA-----TLQPNGGWFTPTSONIGEGHRLTVIATAKGNASPA 1019
QY 378 ---EGIGT-----IYOEITG-----YEPKTAEGTADLPL 409
DB 1020 GNESPESISPTLIDTQADAPQIVSAATYGEVGLLANGSTTNORMTFLSGT----- 1073
QY 410 ALKDIABEKSEGLNAGYIYNTGEEQNLGRLKADGEVIEVKTSSINI--SYTG 467
DB 1074 -----GEGALII--TVYNG-----VELATVOYNPGGSW-- 1100
QY 468 YGEHPESFSLAPNQLSOGINTTLLYRTGTQEWEPVRAAGYVNSIKVNTDTP----- 522
DB 1101 ---TYPLTRR-LSEGNILIT--AAATDAAGNSPTSG--VRSVLLDQPPAQODA 1147
QY 523 ---NNVVYTVDN--HEGKSIYRNSF-----VADLNY-- 550
DB 1148 PLISDVAVPYIGINGNAGANDTPTFSGTGEIGSTIILYNNGSEIGRTVYDGNQSWNT 1207
QY 551 -----EHSITVYOFNSDPDEIRTP--VAFALSTGATADYI--SLGVMVAEVPG 596
DB 1208 PAALTPETTYITTYT--ETDIAGNISPPSASVTFLLDTAANPYITFAEDNVEGVDTIYS 1266
QY 597 GSS--NYPVWSK--DVLTLSEGDYTLWYRFSINNOKDEMKKIGSVYVTPPEYTHP 649
DB 1267 GATTDNTPTVIHGTGIDIGVITLYNNS-----SVYGVTV----- 1301
QY 650 LFEVGHNOTSTYTLDMAHNRVLDP--FTLKNLGLPENGELVYVFRQTOSSGSLMAAOET 707
DB 1302 ---DEGTWTLPLV--TSALPDGYVTLTAADAAG-----NSSGVNSFTEPTVT 1346
QY 708 VHIKGEITFYKRYVE-----GPIPDGSIYRATLAFVNGOOLYLKGRNTYVXI 757
DB 1347 VPLQ-----PPVNEIILDVAVPYGTLTDAFTNDRTLTING-----SGENSTYVI 1393

```


Best Local Similarity 17.7%; Pred. No. 0.6;
Matches 146; Conservative 111; Mismatches 242; Indels 327; Gaps 37;

```

OY 221 GSLVGMGSGTGEEMVWIMNP---GNPDLNLTQSGVDVAATLMRVSASVSSEFENG 277
DB 219 GONVGLTNNYDPDY---QPEGKGBEL-QLSGVITVANGTLTLOAIC--DVSILEYD 270
OY 278 GTSYVYVVALRNNFRYKRSLOLH--VRALYTSQEHMDMREGELASGRPYVYAGNOSIG 335
DB 271 STYS-----LRDN-----PLQVEYCVTAAY-----DESE 295
OY 336 HAFVCD--GYASGTHFPMGMGVSNGFKLLSPTSIGIGEGITFIYOEITIGIE 393
DB 296 SSTVCGTLLHAYTDALILENFENGPVNGW-----LVTDADDGFSWGHYL----- 340
OY 394 PAKTPAAGTDALPILAKDIEAEYKESGSLANGY-----SIYNTGEOSNLDCGRLNKA 449
DB 341 -----NAYDAP-----GHNGHCSLSASIVPGIGPTPDNYLTTPKY 378
OY 450 DG-EVIEVKTSSINISWYGEHSEFSIAPNOLSGINTITLLYRTGTQ---W-E 502
DB 379 EGAKRVKVMYSTODAMMAA-----EHYAVMASTTGAVGDEVILFETMTAKPTGAMYER 433
OY 503 PVRHAG--GYVNSIKVNTD-----PNNVVTYDNNKGLS-I 538
DB 434 TTNLPBGTYIAMRHYNCTDIYELKLDITVEGTPASEPEPVYDFVSLIENKGRKRN 493
OY 539 VPNSF-----VADLNSYEHSTITVOGNSDPBEIRTPVAFALSTGATADVIS 586
DB 494 YRNGYEPDKTDKDLQLAGYNTIANGSLVHIQD-----TVLEYIDETYSRDDQYE 547
OY 587 L-----GWMAEYVGGSSNYPVY 604
DB 548 VECVATVAYNDNIESQVCDKLIYDSQSDIIEGFEAGSIPGGMILLIDADGNVNM--- 604
OY 605 WSKDVLTLESGDTLKRRESINNOKEBKIGS-----VSATPEY-THLPFEGV--- 654
DB 605 -----DYPTMTYG---HDEKCIASPSYLPIMIGVLTPTNYLTPLLEAKLY 649
OY 655 ---HNQSTYTLDMANRVLDPDTLKNLGLPENGELVVFROTOSSGSLMAOETVHT 710
DB 650 KVMWSQADAVYSAE--HYAVMST---GTAYE-DFVLLFETMTAKANGAYERTITL 702
OY 711 KOGERT-----YKRPVEGPIDPSYRATL----- 735
DB 703 PACTKIARHYDCTDMFELLDDITYVSTETVPVTDVYVSLIENKGRKLMVYNG 762
DB 736 -----HAFVNGQOOLYLK----- 748
OY 749 -----GKRNVTAKIVNGTAVEAIESESEIRVFPNPARDYVEISAPCIPQETSILF 799
DB 823 DNIESQVCDKLYNTI-----TSLDNIOGDTSLKITPNPASYVVRLEG-LSRSKSTIELY 876
OY 800 DLSGKIVMKNLSAGHRMDVSRPLNGAVILKYDG---YTKRINT 841
DB 877 NALGICILREETHSEKTEIDVSRINDGVYLIKRVGKNKTKTEKVEI 922

```

RESULT 15

08THC9 PRELIMINARY; PRT: 2951 AA.
AC 08THC9;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Cell surface protein.
GN MA4588.
OS Methanosarcina acetivorans.
OC Archaea: Euryarchaeota: Methanococci: Methanosarcinales;
OC Methanosarcinaceae: Methanosarcina.
OX NCBI_TaxID=2214;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C2A / ATCC 35395 / DSM 2834;
 RX MBLINB-21929760; PubMed-11932238;
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smitov S., Ahnood D., Brown A.,
 RA Allen N., Naylor J., Stange-Thomson N., Dearlilano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grhame D.A., Guss A.M.,
 RA Hedderley R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 RT and physiological diversity.";
 RL Genome Res. 12:532-542(2002).
 DR EMBL: AE011180; AAC07927.1; -;
 KW Complete proteome.
 SQ SEQUENCE 2951 AA; 319558 MW; A6301655D222D25 CRC64;

Query Match

3.6%; Score 158.5; DB 17; Length 2951;

Best Local Similarity 18.7%; Pred. No. 3.4; Indels 379; Gaps 50;

Matches 169; Conservative 111; Mismatches 243;

```

OY 41 SLRMGOTAVSDK-----ISIDYVYRQDAERGITSQEGSPAYFYVANRGNEGYALVAD 96
DB 1988 SLPTITGSDPIRKIFSVTDALQGENELGY-----GPSYTNFAN----- 2029
OY 97 DRIPTILAVSPIGRFMDSPDNLRMLOIYDQELIGLISGKAQLNEILLRTE-GVPAEV 155
DB 2030 -----AVLEVYQVYASAEANFTAN-----VTSGNAPLVEFTDSTGPTGS- 2069
OY 156 HALMDGNHRANDMRNNOGYPRMNRKEPLRPNGHATGCVATAAGIMYHSPLOGESS 215
DB 2070 -----WYDFGDKNSTE---ONPTHTYT-AGTYTVTLVYSNSGSDSEK 2112
OY 216 FDY-HAGSLV-----GNMSGTFGE----- 233
DB 2113 TGYTAGSVYLAIVANFSDQTGTAPLSVQTFDESTNPTMTWEGDKTSEGNPTH 2172
OY 234 MIDWT-----NMPGNPDILNITQSCVDATATLMDVASVNSSTYENGSGTYSY 283
DB 2173 TYETITGYVKLTAATYAGS---NPT-IKTD-YITVTNSVAPVASFEDENSG----- 2221
OY 284 VVGLANRNRFRYKSLQLVHATLYTSEQEHMDMIRGELASGRPVYVYAGNNOGIGHAFVCDGY 343
DB 2222 -----RVPTVO-FTDITTGVSYS----- 2239
OY 344 ASDGTFHFVW--GMGVSNN-----GFYKTLTSPSLIGIGEGIGFTIYOELIT 390
DB 2240 -----WMMDFGDSGNQONPTHTYVTEGVSNTVL---TANGPGG-----SNTIT 2281
OY 391 GIEPAKTPADAGTDA---LPIALAKD-----IEAEKSSGSLANGVS----- 429
DB 2282 STEPVVVASAPLSDSYNGGIPLTNYONGVSGDLMDSDSYAMETSQAKAFTLPSYTDIKW 2341
OY 430 -----IYNTGEOSNLDCGRLN-----KADG-----EYIEVTSINISWYGEHPS 474
DB 2342 ARLYVDVYD-GHEMN-----YKGNVEISIDADDSTYELQKNTFTTYSFPEG----- 2391
OY 475 FSLAPNOLSGINTITLLYRTTEQMEVVRHAGGVNSIKVNTDPPNNVVTVDNN-E 533
DB 2392 -GTGYPWLDHLNRYTSQD-----QMMYDL-----TGEISQGTVNVQAITSKIDSND 2438
OY 534 GKLSIYVNSFVADLNSYEHSTITVOFNSSPDEIR-----TPVAFALSTGATFA-- 581
DB 2439 GRVKAM-----TLVAVYDDSDSEVYVWVNGHDTVPNLPTEY-TGSTSG 2483
OY 582 DDVVISLGMVMAEV-----PGSSSNPVVWMSDYLTL-LSGEGY 617
DB 2484 TSTLASGWSANLTAIYLASVDGIYSFGTTLTSGTPGSGSYGDNTW--DVSSMLTAGET 2541

```



```

OY 102 ILATSPICRFDMS-----MDNLRLMLQIYDQETG-----LILSG----- 137
DB 240 MLAFGVGARYIDRFTANLIGAGORFELPDMGLGVNFVDFSSGDNTRLGIGEGWADY 299
OY 138 -KAOLNEELIRTEG-----VPEVHALMDNGHPANDP-----MRNMGCPW- 178
DB 300 FSSVNA-GYFRBSGHESTYKADYDERPANGDFIRENGYLPSPALGARLMEOYGYADV 358
OY 179 ---NKEPLPNCNHAYTGCVAATAAAQIMRYHSPWLOGESFYHAGSLVGN-----WSGT 230
DB 359 ALFNMADKLQSNPGAATVG-----VATPYPLVTMG-VDRHRT-GNEDLLYSMQ 406
OY 231 FEEMTD-WINMPGNDLNLQSOVDAYATLMRDVSAS-----V 268
DB 407 FYOPDKPW-----SQOLEPQYVNELRLTSGSRDYLVORNNNILEYKKODIL 454
OY 269 SMSFENGSGTYSVYVYVGLRNNFRKRSLOLHVRLYTSQE---WMD-MIRGELASGRV 325
DB 455 SLISIPHDINGTE-----HSTOKIOLNVNSKYGLDRIWDDSLRQO---GGQI 499
OY 326 YVAGNNSIGIHAFCVCDYASDGTFFHNMGMGVNSNGFYKLTLLPSTSLIGEGIGFTY 385
DB 500 QHSGSGSVQDYAILPAYVGG-----SNITKVTARAYRNKNSNNVQLTIT 547
OY 386 -----QEIITGIEPAKTPAEA-GTDALPLALKDIEAEKXSESGLVNGYSIYNTGE 435
DB 548 VLSNGVQVGVQVDTFTADKTSKADGTEAITYTA-----TYKNGV 589
OY 436 EOSNLDLGYRLNKADGEVLEVTSSINISIMWYGEHPEPSFLAPNOLSGINTITLLYR 495
DB 590 AQANPVPSFDIYSGDA-TLSARSATTN-----SSGKATVTLKSSK 628
OY 496 TGTGEWEPYRHAOGGYVNSIKVNTTDPNNVYTVDNNEGLS-IVPNSFVADLNSYENST 554
DB 629 PG-----QVYVSAKTAEMTS-ALNA---NAVIFVQDKTAKITEIKADKTTAKADSDAIT 679
OY 555 IIVQNSDSDPEIRTPVAFALSTGATADVISLGMVMAEVPGSSNYFPVNSKDVLTLE 614
DB 680 YVAKWKNNKQPPANNSVTFSTNG-----DL-----GNSNTQI---VETDKD 719
OY 615 GDYTLMYRFSINNOXDEMKIGSVS---VKTPEYTHPLPEVGHNOTSTYTLDMANRVL 672
DB 720 GRATYKLTLSGVAGNMAIVSAKSEVTEVAPAKFPFSVLSDNN----- 763
OY 673 DDTLKNLGLPENGELVYVFRQTOSSGSLMAAOEYVHIKOGETFYKPYVEGPIDGSR 732
DB 764 ---VNIIGTSANGAL-----PNITW-----LRYGO---FKLAKG--GDGKYQ 797
OY 733 -----ATLHAFVNGOOLYLKGRNTYKLVNG---TAVEAIESSEI-----RV 774
DB 798 WRSOPSVASVDA-LIGRVTLLKKGTGK---IEVSGDKOTATVYTTINTPKIITVETQNV 854
OY 775 FPNPARDVVEISAPCIPOETSIIPLDLSKIVKNSLSAGHRMDVSRPLNGAYILKYDG 834
DB 855 IYSVAEATCSTNSGLPSTST-ELKDYVNNKMGAAANSYEGYKKNITTAQTQADKOSG 913
OY 835 YTTKINIV 842
DB 914 WTSTFDIV 921

```

RESULT 18

O8TP22 PRELIMINARY: PRT: 1075 AA.

AC O8TP22;

DT 01-JUN-2002 (TREMBLrel. 21, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Surface antigen gene.

GN MA1761.

OS Methanosarcina acetivorans.

OC Archaea: Euryarchaeota: Methanococci: Methanosarcinales: Methanosarcinaceae; Methanosarcina.

```

OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Anoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell R.F., Jing H., Macario A.J., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:532-542(2002).
DR EMBL: AE010852; AAM05167.1; -.
KW complete proteome.
SQ
SEQUENCE 1075 AA: 114778 MW: BBE122C2BA209CAB CRC64;

Query Match 3.5%; Score 155; DB 17; Length 1075;
Best Local Similarity 18.9%; Pred. No. 1.2;
Matches 202; Conservative. 129; Mismatches 349; Indels 390; Gaps 57;

OY 1 MKSFLL-----AVIMFGIANGGSHAPYKKEKALSIARLALQVSLRMQGTAVSDKISD 56
DB 1 MKRIFLICIALDMCTGLA-----SAVV---ETVEPVANFTANTTSGPADLVQFTDISTN 54
OY 57 YVYRQDAERGITTSEEGSPAYFY-----VANRGNNEGYA---LVAADRIPPI 102
DB 55 ATGWAMDFGDGNNSTPEQ-NPVHTYSPGTYTSLVSNAGSDSEKTEYISVLELP-- 111
OY 103 LAYSPGRDMD-----SMDNLRLMLQIYD-----QEIGLIIS 136
DB 112 --LAPVANSADITSGVGPLYVHFDTSSSGSPDT---WEMDDNDGDVDSTQNPQFIT 166
OY 137 GRAOLNEELIRTEGVPAEYVHALMDNGHFANDPMRNNQYR-----WNKEPLLPNGNHAY 191
DB 167 GEGTYSVKLTLYSSAGSDSEKTDHTTYV-----YPSDFTANTYEGVPL-TYEF 216
OY 192 TGCVATAAQAQIMRYHSPWLOGESFYHAGSLVGNMSTGTEGMYDINPGNPD---D 247
DB 217 TGNVSGLAPEYWNMMAMVDVGDGYDSENF---NITHYTB-----PGYDVIYVD 266
OY 248 NLQSOVDAYATLMRDV---SASY-----SMSFYENGSGTYSYVYVGLRNNFRK 295
DB 267 GMTQVWKAGYITVLEIPVANSADYVTSMEPLTVNFTDQSTGVSSY----- 314
OY 296 RSLQLHVRALYTSQEMHDIRGELASGRVYVYAGNNSIGIHAFCVCDYASDGTFFHNMGM 355
DB 315 -----SMDPDNDCTVDS-----TEQNPSSH-----YTSAGNYVNL-- 345
OY 336 GGVSNGEFYKLTLLSPTSLGIGEGIGFTYOEIITGIEPAKTPADAGTALPLAKD-- 413
DB 346 -TVSNA-----GGSD-----SELKTDYITVYQAGVATRLISGLVTV 384
OY 414 -IEAEYKSESGLVNGYSIYNTGEEQ-SNDDLGRLNKADGEYIEYKTSINISWGYGEH 471
DB 385 PASAFAARETNPKVLANVQNTGATLNSINIAVAVSDVSGTVPDVTTI-ASLEGDAK- 442
OY 472 PPSFLAPNOLSGINTITLLYRRTGTQWEP-VHRAOGGYVNSIKVNTTDPNNVYVVD 530
DB 443 -----TYTTLI-----DPTIYDLBGGYTYTAAV---VPPDNLIAETD 476
OY 531 -NNEGKLSIVPNSFVADLNSYE-----HSTIVQFNSD-----SPDEIRTPVA 572
DB 477 ETNNKSSAAKN---VRYNGYKGIYWEGGSNITTKHTFDLQGLLXSTQPDASAYQPVG 533
OY 573 FALSTGA-TADYV-----ISLGWMALEVGGSSNRYVWV-----SKDV 609

```



```

Db 534 WESRTEWTASDLPEPDGSTEIEKAFLYVANNM--DQTPGC---YP--WLINENGLNDLN 586
Qy 610 LITSEGGYTLM-----YRFSINNQKDEMKKIGSVSKTPEYTHLPFEVGHN 656
Db 587 GNISTGNGTLYIRDMNSNGAAYNREYGLCYVDYDVKFSAGNSLWTP-----VGEN 637
Qy 657 QTSYV--TLDMAH-----NRVLPDFTLKLGLP----- 682
Db 638 KNALPSTLVVYVYONDNETRKQIFINECEDELGSASSYGTPEATATAYFTGMSIDYE 697
Qy 683 -----ENGELV-----VVFROT----- 694
Db 698 KVTNAMLVSPAGSAPDEGNLLTNGNIVATNANOGSSNGSPLEFADATNINTVGNBAGI 757
Qy 695 -QSSSGSLMAAOEYVHKQGETFYKPYVE-----GPIP-----DGSYRATLHA 737
Db 758 QSTISGGADLQQLVVEYESAPADPADTAPFTGSDAPLANFTDTSSTGSPTSMSWD 817
Qy 738 FVNG-----QOOLY-LGKRNRYVYKIVGTAVEAIESESSEIRFPPNPARDYVEISAPCIP 791
Db 818 FCGDANSTAGNPLHTYSAGEYNT---VNLTVENDAGSOFELKT-----DYIAVSE--AS 866
Qy 792 QETSIIIFDLGSKIVMKNSLSAGHRDVSRLPENGAYILKYDGYTTTINI 841
Db 867 GSTVTLVYFDEPSSSVSEN--ESTEISYVANSNFPSCG-----LSGYNLVTLL 909

RESULT 19
ID 0979C6 PRELIMINARY; PRT; 2076 AA.
AC 0979C6;
DT 01-OCT-2001 (TREMblrel. 18, Created)
DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Hypothetical protein TV1235.
GN TV1235 OR TVG1271204.
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmatales; Thermoplasma.
OX NCBI_TaxID=50339;
RA [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GS1 / DSM 4299 / JCM 9571;
RX MEDLINE=20570466; PubMed=11121031;
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA Kawashima T., Yamamoto Y., Watanabe K., Yamashita M., Kanehori K., Kawamoto T.,
RA Nunoshiba T., Yamamoto Y., Atamaki H., Makino K., Suzuki M.,
RT "Archaeal adaptation to higher temperatures revealed by genomic
sequence of Thermoplasma volcanium."
Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
EMBL: AP000995; BAB60377.1;
InterPro: IPR003674; Orfase-STT3.
DR Pfam: PF02516; STT3; 1.
KM Hypothetical protein: Complete proteome.
SQ SEQUENCE 2076 AA; 228297 MW; AF26CDAFDBE638A5 CRC64;

Query Match 3.5%; Score 154; DB 17; Length 2076;
Best Local Similarity 20.8%; Pred. NO.3.7;
Matches 170; Conservative 116; Mismatches 308; Indels 222; Gaps 46;

```

```

Db 1196 SRIFFYDSDQYNNATVYN-----VSNG--KFSVIVPGVYTIYSGKG--TAYRTIEIT 1247
Qy 390 --TGIEPAKIPA--EAGDAPILALKDIEAEYKESEGLN-----VG- 427
Db 1248 ENQOLNNGSPAYSVSLDS--YVEKAVSSQSYIYVQNGLYGYPSSNSTVHERLPVGI 1305
Qy 428 YSIYNT-----GGEOSNLDGYRLNKADGVEIEKTSINISWYGEHPEPSLA 478
Db 1306 YRISTATITAGLSTAGLVLNLSKNY-----SNTVELSNASIS-----ITALY 1348
Qy 479 PNOLOSINT-----ITLLYRTGTQEWEPVHAAGGYNSI--KYNTDPPNNVYTV 530
Db 1349 NKRISNENFTVDAGSHGIVLLSGNVILLEVPEYRS--GYSNVYPLSTYSPATVLSPY 1406
Qy 531 NNEGKLSI-----VPNSFVADLSYESTITVQFN-----SSPD 565
Db 1407 YLEENISLMAKPITLGMKPVYIEASFELYNKS--LKPEFNGYIDLYGLYNTLVNLSGRAY 1465
Qy 566 EIRTPVAFALS-----TGAT--ADDVLSLGMVAEVPDGSNYPVMSKDLT 611
Db 1466 GYMPGIIYLSINMGNTNLIYNDAITPSSSNLIDVG--IYANFTSDISGVRL--TKGSG 1523
Qy 612 LS-----EGDTLMYRESINNQKDEMKKIGSVKPTETHPLFEVGHNOTSTYTLDMA 666
Db 1524 LSPGPVPIGNITYY-----AKGSKVSVNEIATNTNIVYNSISLANGYLVNTNSLNV 1577
Qy 667 HNRVLPDFTLKLNLPLPENGELV-----VVFROTQSS-----GSLMAOER--- 707
Db 1578 GNYLVKG--SYFVIPGSGFVLDPAGNYIYSFNSRFSSTGYVYGASLILNRNLTLS 1634
Qy 708 VHIQGETFYVKKPVVEGPIPDGSRATLHAFAVNGOOLYLKGRNYT-----VKIVNGTA 762
Db 1635 VVMSKELFANTTI-----LTMSPYS--YVVLNSGSIYAGRANSTCVASFMLPTGFT 1667
Qy 763 VEAISSSEIRFPPNPARDYVEISAPCIPQFTSIILED-----LSGKIYKRN 809
Db 1688 VYSISTNKKMAF-----NYLIN--PFTSYVFNSSLSNSYPVYIYTLANGPEPYLN 1738
Qy 810 -SLAGHRMDYSR-----LPNGAYILKYDGYTTK 838
Db 1739 VVISLGSSELLVNSSVNTIYLPNTGYTFGSSISTTR 1774

RESULT 20
ID 08TUS8 PRELIMINARY; PRT; 2523 AA.
AC 08TUS8;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Cell surface protein.
GN MA3699.
OS Methanococcinae acetiivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanococcinales;
OC Methanococcinae; Methanococcinae.
OX NCBI_TaxID=2214;
RA [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Fitzhugh J.E., Nusbbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
RA Allen N., Naylor J., Strange-Thomann N., Deatellano K., Johnson R.,
RA Linhorst L., McEwan P., McKernan K., Jalinas J., Tirrell A., Ye W.,
RA Zinner A., Barber R.D., Cann I., Graham D.E., Graham D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetler H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayah L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanococcina acetiivorans reveals extensive metabolic

```

RT and physiological diversity.
RL Genome Res. 12:532-542(2002).
DR EMBL/ AE011080; AA07054.1; -
KM Complete proteome.
SQ SEQUENCE 2523 AA; 264288 MW; B3F4DCD961126C5A CRC64;

Query Match 3.4%; Score 151.5; DB 17; Length 2523;

Best Local Similarity 19.5%; Pred. No. 6.9;
Matches 172; Conservative 84; Mismatches 285; Indels 339; Gaps 43;

```
QY 89 GYALVAADDRIPITLAVSPIGRFMDSPDLKMLQIYDQIGLISGKAQLEELRT 148
DB 1025 GYINVTNDSSGVLT---APISTTDASSGRVPTVQFTDTTGNVSTWSMDRFGDETSD 1080
QY 149 EGVAPEYHALMDNGHFANDPMRNOCYPMNKEPL---PGNHAYTCVATAAQAQIMRY 205
DB 1081 EQNP---VHTYVTEG-----THVTLTATGCGTDTATSTMEVAA----- 1117
QY 206 HSNPLQEGSF-----YHAGSLVGNMSTFGEMYPDWINMGNPDLDNLTSQ 253
DB 1118 ---PL-NTGSYDGGSPLLTVQEGTVSGGLMHGSMWPGSGYGEQFTLPPTET- KMARLY 1172
QY 254 VDAVATLMRDV-SASVSMSTFYENGSGTYSTYVVGALNNFRYKSLQHLVRL--YTSQE 310
DB 1173 ADVYCGLMTDIKSFSMTIDIDGGDGIX-----ELQKHDAASIDYNTPT 1216
QY 311 -----WHDN-----IRGELASGR-----PYVAGNNOISIG 335
DB 1217 AWYNDHVRYTSDYFMWYDLTDNITGNEVNIKATGGSDRIKHVTLVAYADGDDDET- 1275
QY 336 HAFVCGGYASDGTFFHN---WGMGVS-----NGFYKLTLLSPSLGIG 376
DB 1276 HYVYNOGHDTANSGYSTGVTGTSALSDPAVNLSTYCAVNGAYTYNGESLSGGPQ 1335
QY 377 GEGIGF---TIYOEITIGIEPAKTPAEAGTDALPILAKDIEAEKSESG----- 423
DB 1336 GEYFGYNDMNVTEYFNSGEDSTVYVAPTGSYKIQIAL- LTRASAGSSGGDGPVADFTA 1393
QY 424 -----LNVGYSIYNTGEEQSNL-DLG-----YRLN---KADG 451
DB 1394 NNTSGGIPLVSTFSTDESTGSPASMLWDFDGNSTEQNPVHTYSTEGTYSVNLVTINEDG 1453
QY 452 EVLEVNTSSINISWYTGHEHESFSLAPNLSOG--INTI-----TLLYRR 495
DB 1454 SDELKTDTYITVQAG-----QVATNDLSISGLVNTVPASAVFARETNPVKVLNVQN 1505
QY 496 TGTQOME-----PVRHAGGYVNSIK 516
DB 1506 TGTATLSNISIAVYASDVSSGIVPDTTIIASLAGDAKTTVLLIDPTIRDEGGVYVYA 1565
QY 517 VNTTDBNNVVYVD--NNEGKLSIVNSFYADLNSY-----HSTIVQFN 560
DB 1566 V--VDDPNLIAETDETNKNSAKPLRY---NGYKKGIIWEGSNITTRHTFDLQGN 1619
QY 561 ---SDSPDEIRTPYAFALSTGA-TADDV-----ISLGWMAAEVPGSSNP 602
DB 1620 LLYSTQPDAYQVPGWESRTETWTASDLPVPGSTIEKAFLVAYANW--DOTPG--Y 1674
QY 603 VVW-----SKDVLTSRGDYTLW-----YRFSINNOKDEMKKIGSYSVK 641
DB 1675 --WLNINFGNALDNGNISTGNGTLYRDMNSNGCAVADYEGLCVYVNTDKFSSAGNSLV 1732
QY 642 TPLEYTHPLFEVGHNOTSTYLLDMAHNRVLPFTLKNLGLPENGELVVVFRQOSSGSL 701
DB 1733 TP-----VGENKNALYP-----STLVVYVGNNDNETRKQI 1761
QY 702 MAAQETVHIKOGTFYKPYVEGPIPDGSTRATLHAFVNG 741
DB 1762 FINEEC-----DELAYSLTGYGTPE---EATAYAPFTG 1792
```

Search completed: May 29, 2003, 13:26:53
Job time : 52 secs